

Systematischer Review Internationaler Evaluationen von Mammographie-Screening- Programmen

Strategien zur Kontrolle von Bias

22.03.2010

Marcial Velasco Garrido, MPH

Dr. med. Annette Zentner, MPH

Prof. Dr. med. Reinhard Busse, MPH, FFPH

Fachgebiet Management im Gesundheitswesen

Technische Universität Berlin

Dr. P.H. Abdel Moniem Mukhtar, MPH, BDS

Lena Burbach, Dipl.-Ges.-Ök

Dr. rer. pol. Dominik Graf von Stillfried

Zentralinstitut für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland

Inhaltsverzeichnis

Abbildungsverzeichnis	3
Tabellenverzeichnis	3
Auftrag	4
1. Einführung	5
1.1 <i>Lead-Time-Bias</i>	5
1.2 <i>Length-Time-Bias</i>	6
1.3 <i>Overdetection- bzw. Over-diagnosis-Bias</i>	6
1.4 <i>Self-Selection-Bias</i>	7
1.5 <i>Andere</i>	7
2. Methodik	7
3. Ergebnisse der Extraktion	9
3.1 <i>Studien-Pool</i>	9
3.2 <i>Allgemeine Charakteristika der untersuchten Screening-Programme</i>	10
3.3 <i>Studientypen und Datenquellen</i>	11
3.2.1 <i>Fall-Kontroll-Studien</i>	11
3.2.2 <i>Kohortenstudien</i>	14
3.2.3 <i>Ökologische Studien</i>	16
3.2.4 <i>(Quasi) Randomisiert</i>	17
3.2.5 <i>Fall-Kohorte</i>	17
3.3 <i>Endpunkte</i>	17
3.4 <i>Berichtete Strategien zum Umgang mit Bias</i>	19
3.4.1 <i>Lead-Time-Bias</i>	19
3.4.2 <i>Length-Time- / Over-Diagnosis-Bias</i>	23
3.4.3 <i>Self-Selection-Bias</i>	25
3.4.4 <i>Verzerrungen im Zusammenhang mit dem Diagnosezeitpunkt und der Exposition</i>	28
3.4.5 <i>Weitere Aspekte</i>	32
3.5 <i>Ergebnisse der Studien</i>	33
4. Zusammenfassende Diskussion	35
5. Extraktionstabellen	38
5.1 <i>Charakteristika der Screening-Programme</i>	39
5.2 <i>Charakteristika der Studien</i>	42
5.3 <i>Ergebnisse der Studien</i>	59
6. Literaturliste	61
6.1 <i>Übermittelte Studien</i>	61
6.1.1 <i>Systematische Übersichten</i>	61
6.1.2 <i>Primärliteratur</i>	61
6.2 <i>Weitere Literatur</i>	63
7. Anhang	65
<i>Recherche Strategie</i>	65
<i>Systematischer Review von Schopper & Wolf</i>	66

Abbildungsverzeichnis

Abbildung 1. Lead-Time-Bias (Quelle: eigene Darstellung nach 57).....	5
Abbildung 2. Flow-Chart der Selektion der Literatur (Quelle: eigene Darstellung).....	9
Abbildung 3. Fall-Kontroll-Studien zum Mammographie-Screening: Definition des Expositionszeitraums (Quelle: eigene Darstellung).....	14
Abbildung 4. „Kontrolliertes Vorher-Nachher“-Design am Beispiel Alterskohorte 50-69 J. (Quelle: eigene Darstellung).....	16
Abbildung 5. Fall-Kontroll-Studie mit Verschiebung des Pseudodiagnosedatums der Kontrollen (Quelle: eigene Darstellung).....	20
Abbildung 6. Umgang mit Lead-Time in Evaluationen des schwedischen Screenings (Bsp. für Screening Alter 50-69 J) (Quelle: eigene Darstellung).....	21
Abbildung 7. Umgang mit Lead-Time-Bias Studie von Tabar 2001 (Quelle: eigene Darstellung).....	23
Abbildung 8. <i>Time of Diagnosis Bias</i> in Fall-Kontroll-Studien (Quelle: eigene Darstellung).....	28
Abbildung 9. Veränderung der Ergebnisse je nach Definition der Exposition (grüne Pfeile) bzw. nach Korrektur für Self-Selection (rote Pfeile) (Quelle: eigene Darstellung mit Angaben der Studien).....	34

Tabellenverzeichnis

Tabelle 1. Studien- und Publikations-Pool.....	10
Tabelle 2. Definition von „Fall“ und „Kontrolle“ in Fall-Kontroll-Studien.	13
Tabelle 3. Unterschied zwischen Inzidenzbasierter Mortalität und Todesfällen unabhängig von Diagnosezeitpunkt	24
Tabelle 4. Intention-to-screen (ITS) Analyse vs. „per Protokoll“ (pP) Analyse (tatsächlich gescreente).....	25
Tabelle 5. Korrektur für Self-Selection-Bias (Beispiele).....	26
Tabelle 6. Self-Selection – Definition der Exposition und Korrektur.....	27
Tabelle 7. Vergleich RR mit und ohne Berücksichtigung der Todesfälle mit BK-Diagnose vor Einführung des Programms.....	29
Tabelle 8. Korrektur für Setup-/Inclusion-Bias (Beispiele).....	30
Tabelle 9. <i>Exposure Opportunity Bias</i> (Beispiele).....	31
Tabelle 10. Programm Charakteristika.....	39
Tabelle 11. Studiendesign, Datenerhebung.....	42
Tabelle 12. Umgang mit Bias.....	47
Tabelle 13. Analyse.....	54
Tabelle 14. Ergebnisse Brustkrebsmortalität.....	59
Tabelle 15. Systematische Übersichtsarbeit Schooper & Wolf 2009.....	66

Auftrag

Das Zentralinstitut für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland beauftragte das Fachgebiet Management im Gesundheitswesen mit der Extraktion und Zusammenfassung der unter Abschnitt 6 gelisteten Studien zur Evaluation von Mammographie-Screening-Programmen.

Ziel der Extraktion und Zusammenfassung dieser Studien sollte sein, sich einen Überblick über die in solchen Evaluationen verwendeten Studiendesigns, insbesondere über den Umgang mit den verschiedenen Biasarten (d.h. Verzerrungspotentialen) zu verschaffen. Es ist nicht Ziel dieses Auftrags eine Bewertung der Effekte der in den Publikationen beschriebenen Mammographie-Screening-Programmen¹ an sich, sondern eine Zusammenfassung der Evaluationsmethoden zu liefern.

Die für diesen Gutachten zur Verfügung gestellten Publikationen wurden vom Zentralinstitut für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland nach einer systematischen Literaturrecherche und Selektion identifiziert und dem Fachgebiet Management im Gesundheitswesen ktion übermittelt.

¹ Beim vorliegenden Gutachten handelt es sich um einen „*rapid review*“; der vom Auftraggeber vorgegebene Zeitraum für die Durchführung der Extraktion und Zusammenfassung erstreckte sich vom 01.12.2009 bis zum 31.01.2010.

1. Einführung

Im Allgemeinen ist zu erwarten, dass Personen, deren Erkrankung durch Früherkennungsprogramme identifiziert wurde, eine längere Überlebenszeit aufweisen als Personen, bei denen die Erkrankung aufgrund von Symptomen diagnostiziert wurde, wobei dies durch verschiedene Biasarten erklärt werden kann (52). Bei der Evaluation von Screening-Programmen müssen deshalb verschiedene Verzerrungsmöglichkeiten berücksichtigt werden, da es anderenfalls zu einer Überschätzung der Effekte eines Screening-Programms kommen wird.

Die einzige umfassende Strategie, Verzerrungen der Ergebnisse durch die am häufigsten beschriebenen Biasarten (*lead-time*, *length-time*, *overdiagnosis* und *self-selection bias*) wirksam zu minimieren, ist die Durchführung von randomisierten Studien unter Verwendung von Mortalität (z.B. Todesfälle pro Personenjahre) als Outcomeparameter (52,53). Aber auch bei nicht-randomisierten Studien ist die Analyse der erkrankungsspezifischen Mortalität weniger anfällig für Verzerrungen als die Analyse der Überlebenszeit.

1.1 Lead-Time-Bias

Unter Lead-Time versteht man die Zeit, die bei der Diagnose einer Erkrankung durch Screening gewonnen wird (Abbildung 1). Damit ein Screening-Programm wirksam sein kann, muss ein Vorteil aus dieser Vorverlagerung der Diagnose entstehen (53), d.h. die Gesamtzeit bis zum Tod muss sich tatsächlich verlängern, indem die Diagnose zu einem Zeitpunkt erfolgt, in dem die Erkrankung noch in ihrer präklinischen Phase und damit wirksamer therapierbar ist. Sollte jedoch lediglich die Diagnose durch Screening früher erfolgen, aber keine wirksamere Therapie möglich sein, wäre die längere Zeit zwischen Diagnosestellung und Tod kein Vorteil, sondern lediglich dem Lead-Time-Bias zuzuschreiben.

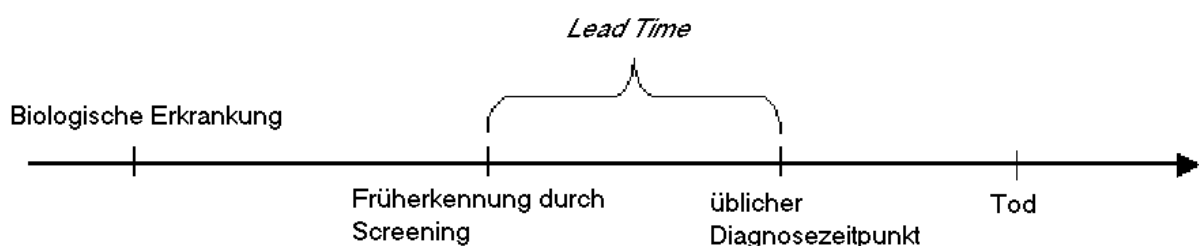


Abbildung 1. Lead-Time-Bias (Quelle: eigene Darstellung nach 57)

Die Lead-Time variiert von Fall zu Fall. Bei manchen Frauen wird durch das Screening die Diagnose um Monate oder gar Jahre vorverlagert, bei anderen wäre die Diagnose aufgrund der klinischen Manifestationen nur wenige Tage später gestellt worden.

Lead-Time muss bei der Ermittlung von Überlebenszeiten berücksichtigt werden. Durch die Lead-Time werden allein durch die Tatsache, dass die Diagnose zeitlich vorverlagert wurde, längere

Überlebenszeiten mit der Diagnose „Brustkrebs“ beobachtet, ohne dass es unbedingt einen wirklichen Nutzen der Früherkennung gäbe.

Es gibt grundsätzlich zwei Möglichkeiten das Risiko einer Überschätzung der Effektivität von Screening durch Lead-Time-Bias zu minimieren (50):

- Evaluation der Wirksamkeit nicht anhand der Überlebenszeit, sondern anhand der altersspezifischen Mortalität
- Wenn eine Abschätzung der Lead-Time vorliegt, kann dies berücksichtigt werden (z.B. bei einer Lead-Time von einem Jahr kann die Case-Fatality-Ratio über sechs Jahre in der Screeninggruppe mit der über fünf Jahre in der Kontrollgruppe verglichen werden) (50).

1.2 Length-Time-Bias

Schnellwachsende Krebserkrankungen, die meistens eine schlechtere Prognose haben als langsamwachsende, werden eher aufgrund der sich schnell entwickelnden Symptomatik diagnostiziert als durch Screening. Dadurch ist zu erwarten, dass durch Früherkennung überwiegend langsamer wachsende Tumoren identifiziert werden, die meistens eine bessere Prognose haben. Diese ungleiche Verteilung von schnellwachsenden Tumoren mit schlechteren Prognosen und kürzeren Überlebenszeiten nach Diagnose und von langsamer wachsenden Tumoren mit besseren Prognosen und längeren Überlebenszeiten nach Diagnose erklärt einen Teil des Unterschieds in der Überlebenszeit zwischen nicht-gescreenten und gescreenten Frauen. Dieser Unterschied ist nicht auf das Screening-Programm zurückzuführen. Um dieses Problem zu minimieren, wird in randomisierten Studien nicht der Diagnosezeitpunkt, sondern der Randomisierungszeitpunkt berücksichtigt (47). Length-Bias ist zwar zu Beginn des Screening-Programms (d.h. in der ersten Screeningrunde) ausgeprägter, bleibt aber ein Problem über die gesamte Dauer des Screening-Programms (52).

1.3 Overdetection- bzw. Over-diagnosis-Bias

Overdetection bzw. *over-diagnosis* (Überdiagnose) ist die Diagnose bzw. Identifizierung von Tumoren, die sich während des Lebens einer Frau nicht zu klinisch manifesten Krebserkrankungen entwickelt hätten und deshalb nur aufgrund der Früherkennung diagnostiziert wurden (45). Überdiagnose stellt ein bedeutendes Problem von Früherkennungsprogrammen dar. Es führt dazu, dass ein Teil der Frauen sich diagnostischen Verfahren und Krebstherapien (einschließlich ihrer unerwünschten Wirkungen) für eine Erkrankung unterziehen müssen, die nie klinisch manifest geworden wäre und entsprechend keine Behandlung erfordert hätte. Wäre diese Erkrankung unentdeckt und unbehandelt geblieben, hätte sie auch nicht zum Tod geführt. Frauen in dieser Situation haben deshalb keinen Nutzen vom Screening.

Bei der Evaluation von Früherkennungsprogrammen stellt Überdiagnose auch ein Problem dar, da es zu einer Überschätzung der positiven Effekte von Screening führt. Betrachtet man die Überlebenszeit oder die Mortalität unter diagnostizierten Fällen, zählen „überdiagnostizierte“ Fälle, die nur in der Screeninggruppe vorkommen, zu den überlebenden. Das Überleben von

„überdiagnostizierten“ Fällen ist jedoch nicht auf das Screening zurückzuführen, sondern darauf, dass es sich um Tumoren mit sehr geringer Invasivität handelt. Diese Verlängerung der Überlebenszeit wird jedoch der Früherkennung zugeschrieben. Bias durch Überdiagnose entspricht der Extremform des *length bias* (45,52).

Andererseits kann es aber auch sein, dass bei Frauen mit einer Brustkrebsdiagnose allein dadurch, dass die Diagnose vorliegt, die Todesursache eher auf den Brustkrebs zurückgeführt wird. Dadurch könnte eine Verzerrung zuungunsten des Screenings entstehen, da einerseits mehr Brustkrebsfälle in der Screeninggruppe vorliegen und andererseits mehr Todesfälle auf Brustkrebs zurückgeführt würden.

1.4 Self-Selection-Bias

Frauen, die an Screening-Programmen für Brustkrebs teilnehmen, haben i.d.R. einen höheren sozioökonomischen Status als solche, die nicht teilnehmen. Ein höherer sozioökonomischer Status ist mit einem höheren Risiko für die Entwicklung von Brustkrebs verbunden. Allerdings ist das Risiko, an Brustkrebs zu sterben, geringer als bei Frauen mit einem niedrigen sozioökonomischen Status (41). Insgesamt führt diese Konstellation zu einer Überschätzung der Wirksamkeit von Screening-Programmen, weil Frauen mit einem geringeren Mortalitätsrisiko eher am Screening teilnehmen. Auch ohne Screening hätten die Teilnehmerinnen eine geringere Brustkrebssterblichkeit. Dieser Bias ist bei der Evaluation von Screeningprogrammen insbesondere relevant, wenn die Outcomes von Teilnehmerinnen mit den Outcomes von Nicht-Teilnehmerinnen verglichen werden.

1.5 Andere

Nach Morrison ist bei der Berechnung von krankheitsspezifischen Mortalitätsraten und ihrem Vergleich zwischen nicht-gescreenten und gescreenten Gruppen zu beachten, dass Frauen, bei denen bereits Brustkrebs diagnostiziert wurde, nicht am Screening teilnehmen (53). Deshalb ist zu erwarten, dass in der Zielgruppe des Screening-Programms weniger Brustkrebstodesfälle auftreten als in der allgemeinen Bevölkerung, denn in der gescreenten Gruppe gehen die Todesfälle der vor dem Screening erkrankten Frauen nicht ein. Diese Differenz sollte berücksichtigt werden, wenn man die in der Screeninggruppe beobachteten Brustkrebstodesfälle mit den erwarteten vergleicht (53).

2. Methodik

Eine systematische Literaturrecherche und Auswahl der relevanten Publikationen wurde vom Zentralinstitut für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland durchgeführt.

Die Recherche wurde in der Datenbank Medline durch die Plattform Pubmed in September 2009 durchgeführt. Die Recherchestrategie kombinierte Suchwörter für Mammographie, Screening und Brustkrebs (siehe Anhang).

Die Publikationen wurden nach folgende Selektionskriterien von zwei unabhängige Personen ausgewählt:

Einschlusskriterien

E1 Interventionsgruppe schließt Patientinnen ein, die in einem flächendeckenden und qualitätsgesicherten Mammographie-Screening-Programm* teilgenommen haben

E2 Kontrollgruppe schließt Patientinnen ein, die NICHT in einem flächendeckenden und qualitätsgesicherten Mammographie-Screening-Programm² teilgenommen haben

E3 Zielgröße ist Brustkrebsmortalität

E4 Studien-Design ist Beobachtungsstudie

E5 Publikationssprache ist Deutsch oder Englisch

Ausschlusskriterien

A1 Randomisierte kontrollierte klinische Studien (RCTs)

A2 Doppelpublikationen ohne relevante Zusatzinformation

A3 Keine Volltext-Publikation verfügbar

Die Ergebnisse der Selektion wurden dem Fachgebiet Management im Gesundheitswesen der Technischen Universität Berlin für eine systematische Extraktion und Zusammenfassung übermittelt.

Nach den Vorgaben des Zentralinstituts für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland wurden aus den Publikationen Angaben zu den relevantesten Aspekten des Studiendesigns sowie zum Umgang mit verschiedenen Biasarten extrahiert und in Evidenztabelle zusammengefasst. Die Anordnung der Tabellen folgt primär dem Studiendesign und innerhalb des Studiendesigns den Ländern, deren Screening-Programme ausgewertet wurden. Publikationen, bei denen das gleiche Kollektiv untersucht wurde, gingen in den Tabellen als eine Studie ein.

Es wurde eine narrative qualitative Zusammenfassung vorgenommen. Diese befasst sich schwerpunktmäßig mit der Beschreibung der verschiedenen Strategien, die in den Studien verwendet wurden, um mit verschiedenen Verzerrungsquellen umzugehen, bzw. mit den Ansätzen, um die

²lang bestehenden und staatlich finanzierten Mammographie-Screening-Programmen sind hauptsächlich aber nicht ausschließlich in Australien, Großbritannien, Kanada, Dänemark, Finnland, Island, Italien, Niederlande, Spanien und Schweden zu finden.

möglicherweise verzerrten Ergebnisse der Studien zu korrigieren. Die Auswirkung verschiedener Maßnahmen auf die Ergebnisse wird jeweils exemplarisch dargestellt.

3. Ergebnisse der Extraktion

3.1 Studien-Pool

Insgesamt wurden vom Zentralinstitut für die Kassenärztliche Versorgung in der Bundesrepublik Deutschland 44 Publikationsreferenzen übermittelt, die in diesem Gutachten bearbeitet wurden. Dabei handelt es sich um eine systematische Übersicht und 43 Primärpublikationen (siehe unter 6.1. Liste der übermittelten Publikationen).

Die eine systematische Übersicht (1) (siehe Anhang) wurde in dieser Arbeit nicht weiter berücksichtigt, da dort zwar die Ergebnisse der eingeschlossenen Studien nach der Korrektur für Self-Selection-Bias berichtet, nicht jedoch die methodischen Ansätze der Korrektur selbst beschrieben wurden³.

Insgesamt waren 43 Primärpublikationen durch die elektronische Recherche identifiziert worden, und uns übermittelt. Eine Publikation (19) konnte nicht ausgewertet werden, da sie nicht geliefert wurde. Von den 42 verbleibenden Publikationen wurden zwei weitere ausgeschlossen, weil sie nicht über organisierte Mammographie-Screening-Programme sondern von opportunistischem Screening berichteten (18,39). Eine Studie (39) berichtete nur deskriptive Statistiken über die Verbreitung von Mammographie in Katalonien (Spanien).

Insgesamt wurden 40 Publikationen analysiert die von insgesamt 33 verschiedenen Studien aus 8 europäischen Ländern sowie Kanada, Australien, und den USA berichteten. Die meisten Studien stammen aus Schweden (siehe Tabelle 1). Die Publikationen wurden zwischen 1984 und 2009 veröffentlicht, wobei die Hälfte nach 2000 veröffentlicht wurde.

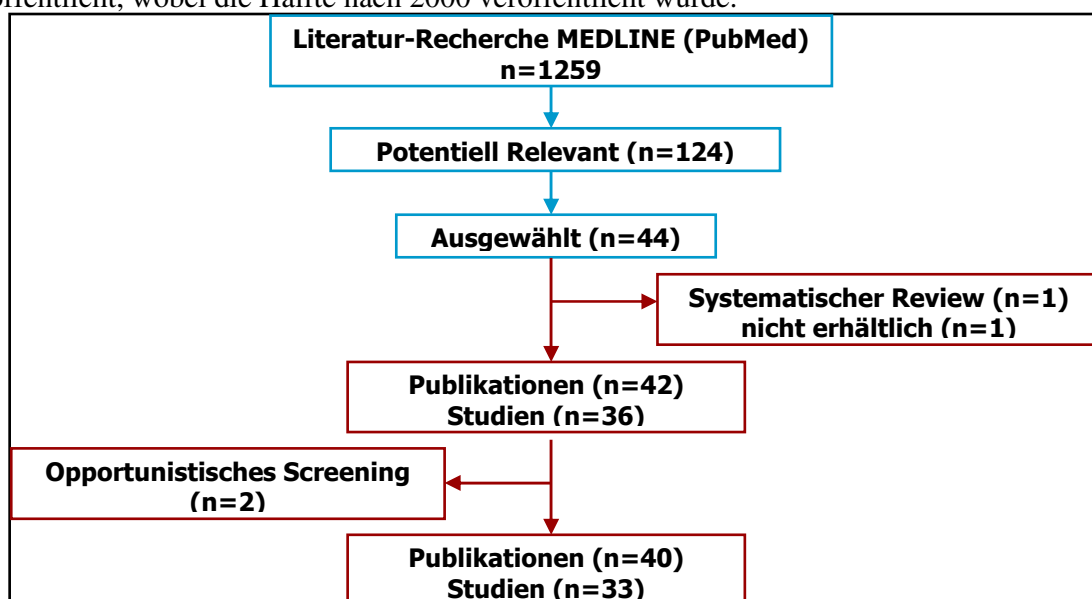


Abbildung 2. Flow-Chart der Selektion der Literatur (Quelle: eigene Darstellung)

³ Die Studien, die in der systematischen Übersicht eingeschlossen wurden, lagen uns vor (siehe Anhang).

Tabelle 1. Studien- und Publikations-Pool

Land	Studien	Publikationen* nach Studiendesign						
		Fall-Kontroll	Kohorte	ökologisch	Quasi-randomisiert	Fall-Kohorte	deskriptiv	?
Australien	2	2	3					
Dänemark	2		(5,6)					
Finnland	4		7,8		(9,10),11			
Großbritannien	4	40,41,42		43				
Island	1	12						
Italien	5	17	15,16	13,14				
Japan	1+1 ^a		4 8					19 ^a
Kanada	1		4					
Niederlande	4	20,(21,22,23,24),26	25					
Schweden	9		27,28,29,30,31,32,(33,34), (35,36)	37				
Spanien	2		38				39	
USA	1					44		
<i>Gesamt</i>	<i>36+1^a</i>	<i>9</i>	<i>19</i>	<i>4</i>	<i>2</i>	<i>1</i>	<i>1</i>	<i>1^a</i>
		<i>42+1^a</i>						
Endgültig eingeschlossen	33	40						

* in Klammern Publikationen, die als einer Studie zugehörig ausgewertet wurden.
^a Diese Studie/ Publikation konnte nicht ausgewertet werden. Durchgestrichen: Publikationen die ausgeschlossen wurden weil sie über opportunistisches Screening berichteten.

3.2 Allgemeine Charakteristika der untersuchten Screening-Programme

Im Folgenden werden die Ergebnisse der Extraktion aus den übermittelten Studien dargestellt. Die Extraktionstabellen befinden sich am Ende des Dokuments in Kapitel 5.

Die Charakteristika der evaluierten Programme sind in Tabelle 10 zusammengefasst. Mehr als die Hälfte der in den Studien evaluierten Screening-Programme wurden in den 1980er Jahren gestartet, und jeweils ca. ein Viertel in den 1970er bzw. den 1990er Jahren. Die Programme waren an unterschiedlichen Altersgruppen gerichtet, wobei unabhängig von der oberen und unteren Altersgrenze die Gruppe von Frauen im Alter zwischen 50 und 69 Jahren in fast allen Programmen eingeladen wurde. Einige Programme (insbesondere aus Schweden) luden Frauen ab ihren 40. Lebensjahr ein. Die obere Altersgrenze wurde häufig bei 74 Jahre festgelegt.

Die meisten Programme hatten einen Screeningintervall von 2 Jahren und (soweit berichtet) fertigten Mammographie-Aufnahmen in zwei Perspektiven, die von zwei Untersuchern ausgewertet wurden.

Aus den Publikationen war meistens nicht ersichtlich, ob es sich dabei um im Rahmen der Mammographie-Screening-Programme geplante Evaluationen oder um ad hoc Auswertungen handelt.

3.3 Studientypen und Datenquellen

Die vorliegenden Studien können nach ihrem Studientyp in folgende Kategorien unterteilt werden: Fall-Kontroll (n=9), Kohorten (n=17), Ökologische (n=4), Quasi-randomisierte (n=2), Fall-Kohorte (n=1). (Übersicht der Studien in Tabelle 1; Details siehe Tabelle 11).

Die Mehrheit der Studien kombinierte Daten aus verschiedenen Quellen, am häufigsten aus Screening-Programm-Registern (die meistens bei der Etablierung der Programme aufgebaut wurden), aus bevölkerungsbezogenen Krebsregistern und aus Todesursachenstatistiken. Insgesamt berichten 31 Studien, Daten aus Mortalitätsregistern bzw. Statistiken herangezogen zu haben. In 30 Studien wurden Daten aus (nationalen oder regionalen) Krebsregistern ausgewertet und in 19 wurden Register der Screening-Programme verwendet. Neun Studien berichten, auch Informationen aus Krankenakten bzw. Krankengeschichte hinzugezogen zu haben, überwiegend um eine höhere Sicherheit beim Assessment der Todesursache zu erreichen oder um Risikofaktoren für multivariaten Analysen zu extrahieren (2,20,21,22,23,24,25,26,41,44). Wie die verschiedenen Datenquellen verlinkt wurden, wird nur in fünf Studien genau beschrieben. Die dänischen Studien verwendeten die persönliche Identifikations-Nummer (5,6), die Studie aus Kanada verwendete eine spezielle Software (*Automatch*) und führte die Verlinkung mittels persönlicher Identifikations-Nummer, Adresse, Name und Geburtsdatum durch (4). Eine australische Studie verlinkte die Datenquellen mittels Geburtsdatum und Postleitzahl (2). In einer Studie aus Holland wurde mittels Name und Geburtsdatum verlinkt (21,22,23,24).

Der Beobachtungs- bzw. Erhebungszeitraum lag zwischen 6 und 29 Jahren. Die Zielbevölkerung (Anzahl der Frauen im Screening-Alter) wurde nur in wenigen Studien angegeben und lag zwischen 13.500 und ca. 1.000.000 Frauen.

3.2.1 Fall-Kontroll-Studien

Bei neun Studien handelt es sich um Fall-Kontroll-Studien, wobei als Fälle immer Frauen definiert wurden, die an Brustkrebs gestorben waren. Die Auswahl von „Tod durch Brustkrebs“ als Definition von Fällen bedeutet, dass die Abschätzung der Exposition aus z.B. Krankenakten oder Registern erfolgen muss. In den Studien wurden dazu u.a. die Dokumentationen bzw. Register der Screening-Programme oder auch klinische Informationsquellen genutzt.

Bei der Definition der Fälle und Kontrollen muss berücksichtigt werden, dass keine Verzerrungen entstehen, z.B. weil die Kontrollen weniger Möglichkeiten hatten gescreent zu werden. Um dies zu gewährleisten, sollen die Kontrollen frei von Brustkrebs zum Zeitpunkt der Diagnose der zugehörigen Fälle sein (2). Die folgende Tabelle 2 gibt einen Einblick in die Definitionen von Fällen und Kontrollen, die in den Studien zur Anwendung kamen. Mit einer Ausnahme (20) waren alle Fall-Kontroll-Studien nach Alter gematcht.

Die Zahl der Fälle (Brustkrebstodesfälle) variierte bei den Fall-Kontroll-Studien zwischen 82 und 1750. Je nach Studie, wurden zwischen 2 und 5 Kontrollen je Fall eingeschlossen.

Tabelle 2. Definition von „Fall“ und „Kontrolle“ in Fall-Kontroll-Studien.

Studie	Definition für Fall	Definition für Kontrolle	Anmerkungen
Roder 2008 ²	Alle an BK verstorbene Frauen im Alter 45-80 zwischen 2002-2005 und deren Diagnose ab 1994 (Jahr der vollen Implementierung des Programms) gestellt wurde	lebten zum Todeszeitpunkt der Fälle und waren BK-frei zum Zeitpunkt der BK-Diagnose der Fälle.	k.A.
Allgood 2008 ⁴⁰	An BK verstorbene Frauen mit Diagnose im Alter von 50-70 J und Todeszeitpunkt ab 6 Jahre nach Implementierung des Programms (ab 1995)	lebten zum Todeszeitpunkt der Fälle und waren BK-frei zum Zeitpunkt der BK-Diagnose der Fälle.	Kontrollen erhalten Datum einer „pseudodiagnose“ = Diagnose-datum der gematchten Fälle
Fielder 2004 ⁴¹	An BK verstorbene Frauen im Alter von 50-75, deren BK-Diagnose nach voller Implementierung des Screening-Programms lag (1991)	lebten zum Todeszeitpunkt der Fälle	Kontrollen erhalten Datum einer „pseudodiagnose“ = Diagnose-datum der gematchten Fälle
Moss 1992 ⁴²	An BK verstorbene Frauen, die in der Studie TEDBC in der Guildford und Stoke Region teilnahmen	lebten zum Todeszeitpunkt der Fälle und waren BK-frei zum Zeitpunkt der BK-Diagnose der Fälle.	Kontrollen waren frei von BK zum Zeitpunkt der BK-Diagnose der Fälle
Gabe 2007 ¹²	An BK verstorbene Frauen, deren BK-Diagnose im Screeningalter (≥ 40 J/ GD ≥ 1919) nicht vor Zeitpunkt der Implementierung des Programms gestellt wurde und deren Todeszeitpunkt nach Ende der 1. Vollscreeningrunde (01/1990) lag	lebten zum Todeszeitpunkt der Fälle	Kontrollen erhalten Datum einer „pseudodiagnose“ = Diagnose-datum der gematchten Fälle
Puliti 2008 ¹⁷	An BK verstorbene Frauen (nach Totenschein) deren Diagnose im Alter von 50-74 gestellt wurde.	lebten zum Todeszeitpunkt der Fälle und waren BK-frei zum Zeitpunkt der BK-Diagnose der Fälle.	Kontrollen erhalten Datum einer „pseudodiagnose“ = Diagnose-datum der gematchten Fälle - 1Jahr und waren frei von BK zum Zeitpunkt der BK-Diagnose der Fälle
Broeders 2002 ²⁰	An BK verstorbene Frauen, deren BK während der jüngsten 10 J des Programms diagnostiziert wurde.	lebten zum Todeszeitpunkt der Fälle in Nijmegen, wurden in der gleichen Screeningrunde wie die Fälle eingeladen, waren frei von BK	k.A.
Collette 1984 ^{21,22,23,24}	An BK verstorbene Frauen, die zwischen 1911 und 1925 geboren wurden und deren BK-Diagnose nach Beginn des Screeningprojekts (1974) gestellt wurde	lebten zum Todeszeitpunkt der Fälle in Utrecht	k.A.
van Dijk ²⁶	An BK verstorbene Frauen, die eine 1. Einladung zum Screening im Alter von ≥ 65 J erhalten hatten.	lebten zum Todeszeitpunkt der Fälle in Nijmegen, wurden in der gleichen Screeningrunde wie die Fälle eingeladen, waren frei von BK	k.A.

BK: Brustkrebs, GD: Geburtsdatum, J: Jahre

Quelle: eigene Darstellung nach Angaben in den Studien

In Fall-Kontroll-Studien muss der Expositionszeitraum definiert werden, also bis zu welchem Zeitpunkt die Screeninggeschichte für Fälle und Kontrollen berücksichtigt wird. In den hier berücksichtigten Studien wurde der Expositionszeitraum für Fälle als die Zeit zwischen der Einführung des Screenings und dem Zeitpunkt der Brustkrebsdiagnose definiert, da Frauen mit Brustkrebs nicht mehr am Screening teilnehmen können.

In den Studien sind die Kontrollen frei von einer Brustkrebsdiagnose zum Diagnosezeitpunkt der Fälle (aber nicht unbedingt zum Zeitpunkt des Todes). Üblicherweise wird dann ein Datum einer „Pseudodiagnose“ gewählt, um den Expositionszeitraum der Kontrollen zu begrenzen: Die Screeninggeschichte der Kontrollen wird beim Zeitpunkt der Brustkrebsdiagnose der gematchten Fälle zensiert (siehe Abbildung 3). Beim Assessment der Screeninggeschichte wird meistens dann ermittelt, ob die Frauen eine Screeningmammographie im fraglichen Zeitraum erhalten haben oder nicht. Dieser Vergleich ist stark durch Self-Selection beeinflusst, was eine Korrektur der Ergebnisse erfordert.

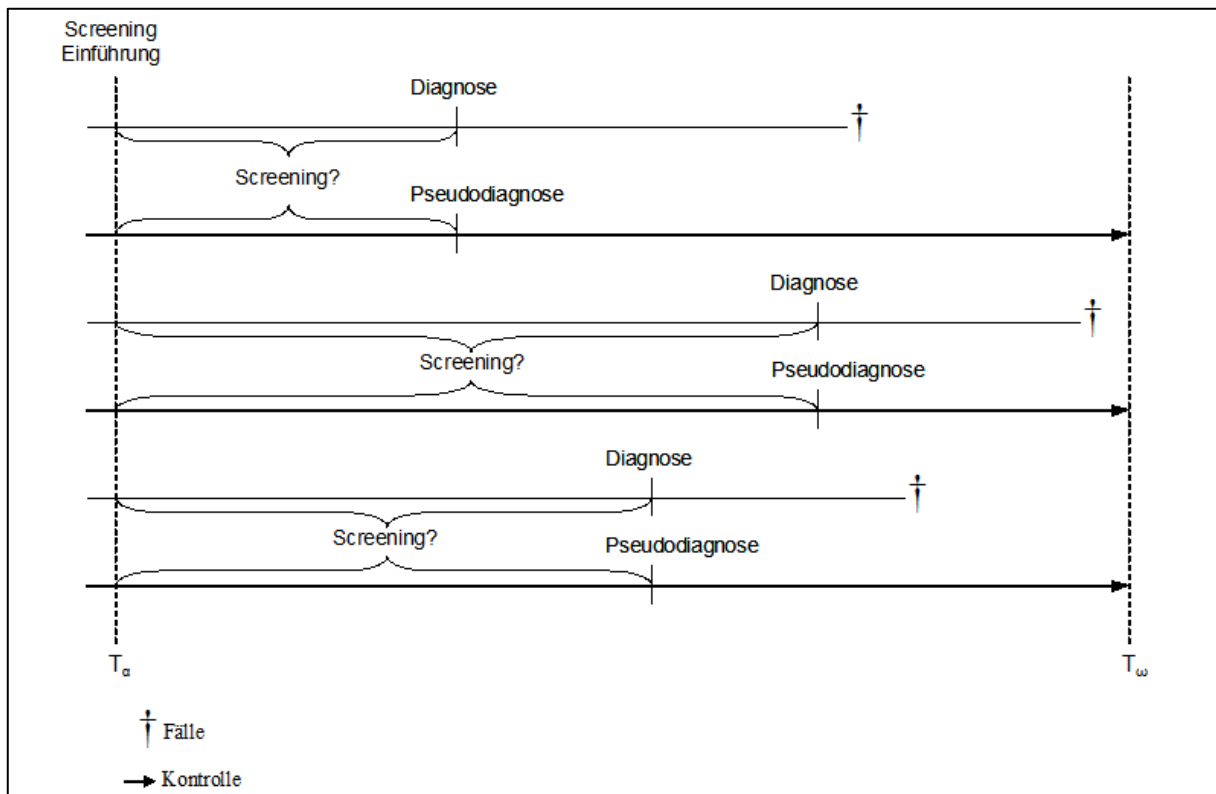


Abbildung 3. Fall-Kontroll-Studien zum Mammographie-Screening: Definition des Expositionszeitraums (Quelle: eigene Darstellung)

3.2.2 Kohortenstudien

Im Allgemeinen definieren sich Kohortenstudien dadurch, dass die Studiengruppen danach gebildet werden, ob eine Exposition vorliegt oder nicht (50).

Bei den Kohorten-Studien variierten die Brustkrebstodesfälle in den exponierten Gruppen (Mammographie-Screening) zwischen 59 und 1182 und bei den nicht exponierten zwischen 71 und 1253. Die Zahl der Brustkrebstodesfälle (Ereigniszahl) war in jeweils ca. 50% der Kohorten-Studien über 1000 und in ca. 25% unter 500. Sofern in den Kohorten-Studien berichtet, variierte die Zahl der exponierten zwischen 60.000 und über 500.000 Frauen, und die der nicht-exponierten zwischen 40.000 und über 500.000 Frauen. Häufiger wurden die Personen-Jahre berichtet, die bei den exponierten zwischen ca. 160.000 und 8.000.000, und bei den nicht exponierten zwischen 155.000 und 21.000.000 lagen.

In den Kohortenstudien zur Evaluationen von Mammographie-Screening wurden grundsätzlich folgende Vergleiche zwischen „exponierten“ und „nicht exponierten“ Gruppen durchgeführt:

(a) einem Mammographie-Screening-Programm exponiert vs. nicht-exponiert / eingeladen vs. nicht eingeladen

Dabei wird die Mortalität in der Gruppe von Frauen, die zum Screening eingeladen wurden, mit der Gruppe von Frauen verglichen, die nicht eingeladen wurden – unabhängig davon, ob bei den eingeladenen Frauen tatsächlich eine Mammographie durchgeführt wurde oder nicht. Dieser Vergleich gibt Auskunft über die Effektivität des Programmes und berücksichtigt per se die Effekte der *Self-Selection*. In den vorliegenden Studien wurden dann die Kohorten geographisch oder zeitlich gebildet.

Zum Beispiel verglich eine Studie aus Holland die relevante Alterskohorte aus einer Stadt mit Screening-Programm (Nijmegen) mit der einer Stadt ohne Screening-Programm (Arnhem) über einen parallelen Beobachtungszeitraum, der mit der Einführung des Screenings in der exponierten Region beginnt⁴ (25). Innerhalb desselben geographischen Raums kann die Kohorte der Nicht-Exponierten aus der Zeit vor und der Exponierten aus der Zeit nach Screeningeinführung gebildet werden. Die Alterskohorte vor Einführung des Screenings wird mit der gleichen Alterskohorte nach Einführung des Screenings verglichen. Dieses Studiendesign könnte als nichtkontrolliertes Vorher-Nachher-Design beschrieben werden. Die Alterskohorte vor der Einführung des Screenings stellt die der Intervention nicht exponierten Frauen dar, während die Alterskohorte nach der Einführung des Screening-Programms die Interventionsgruppe darstellt. Dieses Studiendesign wurde z.B. in den Arbeiten der Arbeitsgruppe um Tabar und Duffy für die Evaluation des Mammographie-Screening-Programms in einigen Regionen Schwedens (27,33,34,35,36), in der Evaluation des italienischen Mammographie-Programms (15,16), in einer finnischen Studie (8) und in einer spanischen Studie (38) verwendet.

Ein häufig verwendetes Kohorten-Studiendesign stellt einer Kombination aus beiden o.g. Ansätzen dar. Dieses Studiendesign könnte als kontrolliertes Vorher-Nachher-Design beschrieben werden und ist in Abbildung 4 dargestellt. Die Outcomes in einer Kohorte von Frauen, die zum Screening eingeladen (bzw. gescreent) wurden, werden mit denen einer Gruppe von Frauen, die nicht eingeladen (bzw. gescreent) wurden, verglichen. Die Exposition erfolgt aufgrund des Wohnorts. Zusätzlich wird sowohl für die Kohorte der exponierten als auch der nicht exponierten Frauen (also in den Regionen mit und ohne Screening) in der Zeit *vor* Einführung des Screenings eine Referenzkohorte (eine vergleichbare Alterskohorte über einen vergleichbaren Beobachtungszeitraum) gebildet. Damit kann zunächst abgeschätzt werden, ob grundsätzliche Unterschiede z.B. in der Mortalität in den beiden Kohorten vorliegen. Die Mortalität während der Studienperiode (also die Periode nach der Einführung des Screenings) kann mit der Mortalität in der Referenzperiode verglichen werden.

Um geographisch bedingte potentielle Risikounterschiede (z.B. im Brustkrebsrisiko – aufgrund unterschiedlicher soziodemographischer Strukturen oder der unterschiedlicher Verteilung von anderen Risikofaktoren – bzw. im Mortalitätsrisiko bei einer Brustkrebsdiagnose, z.B. aufgrund unterschiedlicher Qualität der Versorgung) zu berücksichtigen, können die Ergebnisse mit Hilfe der Referenzperiode adjustiert werden. Die Adjustierung kann z.B. dadurch erfolgen, dass man das relative Risiko (RR) in der Beobachtungsperiode durch das RR in der Referenzperiode teilt

⁴ Screening begann 1975, der Beobachtungszeitraum in der nicht-exponierten Region begann jedoch 1976, was der Vollendung der Hälfte der 1. Screeningrunde entspricht.

(31). Dieses Studiendesign wurde z.B. in der Evaluation der Screening-Programme in Schweden (28,29,30,31,32), Dänemark (5,6) verwendet.

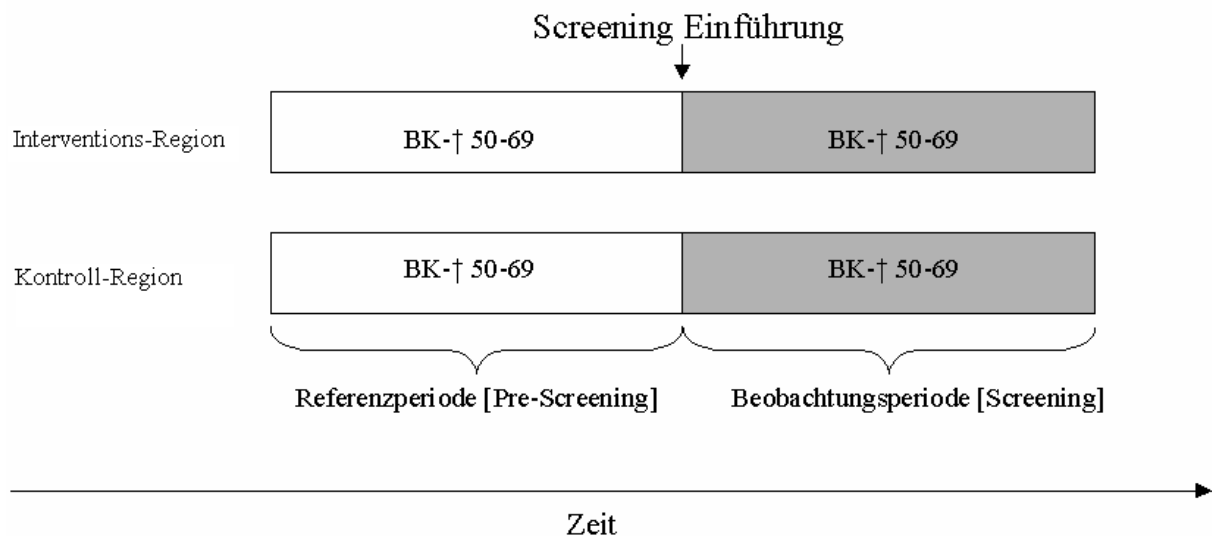


Abbildung 4. „Kontrolliertes Vorher-Nachher“-Design am Beispiel Alterskohorte 50-69 J. (Quelle: eigene Darstellung).

(b) tatsächlich gescreent vs. nicht-gescreent

Dabei gelten als „exponiert“ nur diejenigen Frauen, die zum Screening eingeladen wurden *und* tatsächlich eine Mammographie erhielten. Diese Gruppe kann entweder mit der Gruppe von Frauen verglichen werden, die eingeladen wurden und nicht teilnahmen (also keine Mammographie erhielten) *oder* mit einer Gruppe von Frauen, die nicht eingeladen wurden (und entsprechend nicht teilnahmen). Wie bei den meisten Fall-Kontroll-Studien sind diese Vergleiche stark durch *Self-Selection* beeinflusst, was eine Korrektur der Ergebnisse erfordert. Nur wenige Kohortenstudien definierten die Exposition ausschließlich als die tatsächliche Durchführung einer Mammographie (z.B. 3,4). In manchen Studien, bei denen Ergebnisse für den Vergleich „eingeladen“ vs. „nicht eingeladen“ berichtet wurde, wurde innerhalb der Kohorte, die zum Screening-Programm exponiert war, untersucht, ob es Unterschiede zwischen tatsächlicher Teilnahme und Nicht-Teilnahme gab (z.B. 27,33,34,35,36).

3.2.3 Ökologische Studien

Das Hauptmerkmal von ökologischen Studien besteht darin, dass die Exposition nicht mit den Outcomes auf individueller Ebene verlinkt werden kann (50). Ein Teil der vorliegenden Studien vergleicht die Brustkrebsmortalitätsrate in der Frauenbevölkerung eines Landes oder Region vor und nach Einführung des Screening-Programms (z.B. 13, 43). Studien, die den Zusammenhang zwischen Grad der Einführung/Implementierung des Screening-Programms und Brustkrebsmortalität untersuchen (37) oder zwischen früher und späterer Einführung des Programms (14), folgen auch einem ökologischen Design. Rückschlüsse auf die Effektivität des Screening-

Programms sind jedoch nicht möglich, da unter anderem aus den Daten nicht hervorgeht, welche Frauen tatsächlich dem Programm exponiert waren (also welche Frauen eingeladen wurden).

3.2.4 (Quasi) Randomisiert

Die Autoren der Evaluation des finnischen Screening-Programms beschrieben ihr Studiendesign als randomisiert (9,10,11). In Finnland wurden bei der Einführung des Screening-Programms zunächst nicht alle Frauen der relevanten Altersgruppe eingeladen, sondern es wurde eine Interventionsgruppe und eine Kontrollgruppe gebildet. Die Zuordnung zu den jeweiligen Gruppen erfolgte nach dem Geburtsjahr, wobei Frauen mit geraden Geburtsjahren zum Programm eingeladen wurden und Frauen mit ungeraden Geburtsjahren nicht eingeladen wurden (9,10,11). Aus unserer Sicht kann dieses Design als quasi-randomisiert beschrieben werden.

3.2.5 Fall-Kohorte

Eine Studie wurde von den Autoren als Fall-Kohorte beschrieben. Dabei gehen in die Analyse alle Fälle der Kohorte (also alle Brustkrebstodesfälle) und als Kontrolle eine zufällige Stichprobe aus der Kohorte (Subkohorte) ein, wobei im Gegensatz zu Fall-Kontroll-Studien nicht gematcht wird (44).

3.3 Endpunkte

Durch die Auswahl der Endpunkte für die Evaluation kann im Vorfeld dem Einfluss bestimmter Biasformen Rechnung getragen werden. Wie oben erwähnt (siehe 1.1) kann die Verzerrung durch Lead-Time-Bias reduziert werden, wenn statt der Überlebenszeit die altersspezifische Mortalität betrachtet wird.

In fünf Studien wurden ausschließlich andere Endpunkte als Brustkrebsmortalität verwendet. Eine Studie aus Spanien berichtete ausschließlich Prozessparameter (% Frauen mit Screening, % Frauen mit Biopsie nach Screening) (39). In einer Studie aus Italien wurde ausschließlich das Überleben berichtet (15). Eine Studie aus Australien, berichteten neben dem Überleben das Staging der Tumoren (3) und den Umfang der durchgeführten Therapie bzw. chirurgischen Eingriffs (z.B. konservativ, Mastektomie) (3). Überleben wurde auch in drei weiteren Studien berichtet, allerdings als sekundärer Parameter nach Brustkrebsmortalität (13,21,35,36).

Mit den o.g. Ausnahmen berichten alle vorliegenden Studien über Brustkrebsmortalität als Endpunkt. Allerdings können sich die Studien darin unterscheiden, welche Todesfälle als Fälle eingeschlossen wurden bzw. für die Berechnung der Mortalitätsraten einbezogen wurden. Bei den Fall-Kontroll-Studien wurden nur die Todesfälle aus der Gruppe von Frauen berücksichtigt, deren Brustkrebsdiagnose im Alter der Screeningzielgruppe und nach Einführung des Programms gestellt wurden (siehe Tabelle 2). Auch bei den Kohortenstudien berücksichtigt die Mehrheit der Studien nur die Todesfälle von Frauen mit Brustkrebs, deren Diagnose in der relevanten Altersgruppe (Zielgruppe des Screenings) nach der Einführung des Programms gestellt worden war⁵. Einige Autoren nennen dieses Vorgehen *refined mortality* (10,28,29,30,31,32), andere

⁵ In Tabelle 12 wird der Zeitraum, der mit dem Einführungszeitpunkt des Screenings anfängt und mit dem Ende des Beobachtungszeitraums endet, als „Screening Inzidenzzeitraum“ bezeichnet.

incidence based mortality (7,16,27,33,34) – im Gegensatz zur Gesamt-Brustkrebsmortalität, bei der alle während des Beobachtungszeitraumes aufgetretenen Brustkrebstodesfälle eingehen würden, unabhängig davon wann die Brustkrebsdiagnose gestellt wurde. Mit dieser Beschränkung wird dem Problem Rechnung getragen, dass die Todesfälle, deren Diagnose vor Einführung des Screenings vorlag bzw. bei Frauen, die nicht Ziel der Intervention waren, nicht durch die Intervention hätten beeinflusst werden können. Darüber hinaus scheiden Frauen mit einer Brustkrebsdiagnose für das Screening aus und deshalb könnten ihre Todesfälle nur der nicht exponierten Gruppe zugerechnet werden, was bei Fall-Kontroll-Studien zu Verzerrungen führen kann (siehe auch Seite 28, *time of diagnosis bias*). Die Anwendung dieser *incidence based* bzw. *refined mortality* soll die potentiellen Verzerrungen durch Length- bzw. Over-Diagnosis-Bias reduzieren (16,33) (siehe auch Seite 23).

Eine zweite Frage ist, wie ein Todesfall als Brustkrebstodesfall definiert wird. In Studien, bei denen klinische Daten verwendet werden konnten, kann die Zuschreibung der Todesfälle zur Brustkrebs anhand genauer klinischer Kriterien erfolgen. Zum Beispiel wurde in der Studie aus Nijmegen Brustkrebs als Todesursache akzeptiert, wenn Fernmetastasen identifiziert worden waren und andere Todesursachen ausgeschlossen werden konnten (25). In anderen Studien standen nur Routinedaten (wie z.B. aus Krebsregistern und/oder Todesursachen-Statistik) zur Verfügung, was die Genauigkeit der Zuordnung beeinträchtigen kann. Zwischen verschiedenen Routinedaten-Quellen können zudem Abweichungen auftreten, wie in den Arbeiten der Gruppe um Jonsson für Schweden berichtet wird (28,29,30,31,32). So kann es vorkommen, dass eine Frau, deren Totenschein Brustkrebs als Todesursache aufweist, nicht im Krebsregister als Brustkrebsfall aufgenommen wurde. In Totenscheinen können verschiedene Todesursachen registriert werden. Bei Frauen, die an Brustkrebs erkrankt sind, ist es wahrscheinlich, dass Brustkrebs unter den Todesursachen kodiert wird, auch wenn im Einzelfall der Brustkrebs nicht die unmittelbare Todesursache war. In den Studien der Gruppe um Jonsson wurde deshalb auch die Mortalität mit Brustkrebs als individuelle Todesursache ermittelt. Dabei wurden nur die Todesfälle berücksichtigt, deren Totenscheine Brustkrebs als unmittelbare Todesursache aufweisen (UCD: Underlying Cause of Death). In diesen Studien wurden ähnliche Ergebnisse mit UCD und Exzess-Brustkrebsmortalität (siehe unten) beobachtet, wobei bei älteren Frauen der Unterschied zwischen UCD und Exzess-Brustkrebsmortalität größer war und UCD konservativere Ergebnisse lieferte (28,32).

Die Gruppe um Jonsson berichtet die *Exzess-Brustkrebsmortalität*. Die Betrachtung dieses Parameters soll ein potentieller Bias aufgrund der höheren Anzahl an Brustkrebsdiagnosen bei der Screeninggruppe Rechnung tragen (29,30,31,32). Die Anwendung der Exzess-Mortalität kann deshalb als eine Strategie gegen den Bias durch Überdiagnose verstanden werden. Die Exzess-Mortalität wird als $(O - E)/N$ ermittelt. Die Anzahl von Exzesstodesfällen aufgrund von Brustkrebs ist $(O - E)$, wobei die beobachteten Fällen O die Anzahl aller Todesfälle bei Frauen mit Brustkrebs unabhängig von der Todesursache sind und E die erwartete Anzahl von Todesfällen, wenn es keine Brustkrebsdiagnose gegeben hätte. Die erwartete Anzahl an Todesfällen (E) ergibt sich aus dem Produkt der Personen-Jahren der Brustkrebsfällen und der jährlichen Gesamtmortalität in der Bevölkerung unterteilt in 5-Jahres Altersgruppen und addiert⁶. N ist die gesamte Anzahl an Personenjahren in der Studiengruppe (51).

⁶ \sum (Altersspezifische Mortalitätsrate der Gesamtbevölkerung x Personen-Jahre der Frauen mit Brustkrebs)

Brustkrebsmortalität vs. Gesamtmortalität

Die Anwendung von Brustkrebsmortalität als Outcomeparameter für die Evaluation von Mammographie-Screening-Programmen ist in der Literatur kritisiert worden und manche Autoren plädieren für eine Evaluation auf Basis der Gesamtmortalität (55). In der Evaluation des schwedischen Programms der Two-County-Studie wurde neben der Brustkrebsmortalität auch die Gesamtkrebsmortalität (also einschließlich anderer Tumoren) und die Gesamtmortalität ermittelt, um zu vergleichen in wie weit die Ergebnisse sich unterscheiden. Die Auswahl der Outcomeparameter hatte in diesem Fall keinen Einfluss auf den Ergebnisse: Brustkrebsmortalität RR 0,55 (95%-CI 0,44-0,70), Gesamtkrebsmortalität RR 0,56 (95%-CI 0,44-0,71) und Gesamtmortalität RR 0,56 (95%-CI 0,45-0,70) (35). Eine ähnliche Sensitivitätsanalyse wurde in einer Fall-Kontroll-Studie aus Italien durchgeführt, wobei die Betrachtung aller Todesursachen bei Frauen mit Brustkrebs zu einer konservativeren Schätzung des Effektes führte: Brustkrebsmortalität OR 0,75 (95%-CI 0,62-0,92) vs. Gesamtmortalität OR 0,83 (95%-CI 0,70-0,98) (17). In der Studie aus Dänemark wurde auch der Zusammenhang zwischen Gesamtmortalität und Screening untersucht, wobei das Ziel der Autoren die Validierung des Studiendesigns für die Untersuchung der BK-Mortalität war (5,6).

3.4 Berichtete Strategien zum Umgang mit Bias

Im Folgenden werden die Strategien zusammengefasst, die in den Studien beschrieben wurden, um mit verschiedenen Biasarten umzugehen. Wie in der Einführung beschrieben, sind in Studien zur Evaluation von Screening-Programmen insbesondere Lead-Time-, Length-Time-, Over-Diagnosis- und Self-Selection-Bias relevant. Neben diesen Biasarten wurden in den vorliegenden Studien auch andere Verzerrungspotentiale bzw. andere Probleme beschrieben, die die Validität der Studien beeinträchtigen können. Die Bezeichnungen von ähnlichen Biasformen werden in Studien nicht immer konsistent verwendet. In vielen Studien werden Designcharakteristika beschrieben, ohne diesen explizit die Vermeidung einer bestimmten Biasform zuzuschreiben⁷.

Die folgenden Beschreibungen werden mit Beispielen aus den Studien erläutert und mit ggf. eigenen Abbildungen illustriert, um die Verständlichkeit der Darstellung zu erhöhen.

3.4.1 Lead-Time-Bias

In den hier extrahierten Studien wurde die Problematik des Lead-Time-Bias überwiegend dadurch berücksichtigt, dass die Effektivität der Mammographie-Screening-Programme anhand der altersspezifischen Brustkrebsmortalitätsraten statt anhand von Überlebenszeiten evaluiert wurde. Dementsprechend wurde diese Form von Bias in wenigen Studien explizit genannt. Durch die Auswahl des Endpunkt „Brustkrebsmortalität“ kann erwartet werden, dass der Einfluss von Lead-Time-Bias minimiert wird (50) (siehe auch im Kapitel 1.1).

In einigen Studien wurden jedoch die Effekte vom Screening auf das Überleben der Brustkrebspatientinnen untersucht (3,13,15,21,35,36). In einer von diesen Studien wurde eine Lead-Time von 3 Jahren angenommen und dann die Überlebensraten bei acht Jahren in der

⁷ In Tabellen bzw. Text wurde eine möglichst einheitliche Bezeichnung für ähnliche Verzerrungsquellen verwendet, die jedoch nicht immer der von den Autoren der betreffenden Studien verwendeten entspricht.

Interventionsgruppe mit den Überlebensraten nach fünf Jahren in der Kontrollgruppe (statt 5-Jahresüberlebensraten in beiden Gruppen) verglichen (3).

In der niederländischen Studie DOM wurde eine Analyse des Überlebens durchgeführt, wobei die Überlebenszeit der Frauen mit durch Screening diagnostizierten Brustkrebs um 2,5 Jahre reduziert wurde, wenn er beim ersten Screening identifiziert worden war – und um 1,25 Jahre, wenn er in den nachfolgenden Screeningrunden identifiziert wurde (21).

In einigen Studien wurde der potentielle Einfluss von Lead-Time auch für den Endpunkt Brustkrebsmortalität berücksichtigt bzw. die Ergebnisse (Mortalitätsrisiko) für Lead-Time korrigiert.

In der IMPACT Studie (17) – einer Fall-Kontroll-Studie – wurde der Zeitpunkt der Pseudodiagnose der Kontrollen (siehe Abbildung 3) nicht mit dem Zeitpunkt der Diagnose der Fälle gleichgesetzt, sondern um ein Jahr verschoben (siehe Abbildung 5). Damit sollte die Verzerrung durch Lead-Time reduziert werden (17). Darüber hinaus wurden in dieser Studie Sensitivitätsanalysen mit anderen „Verschiebungszeiträumen“ durchgeführt.

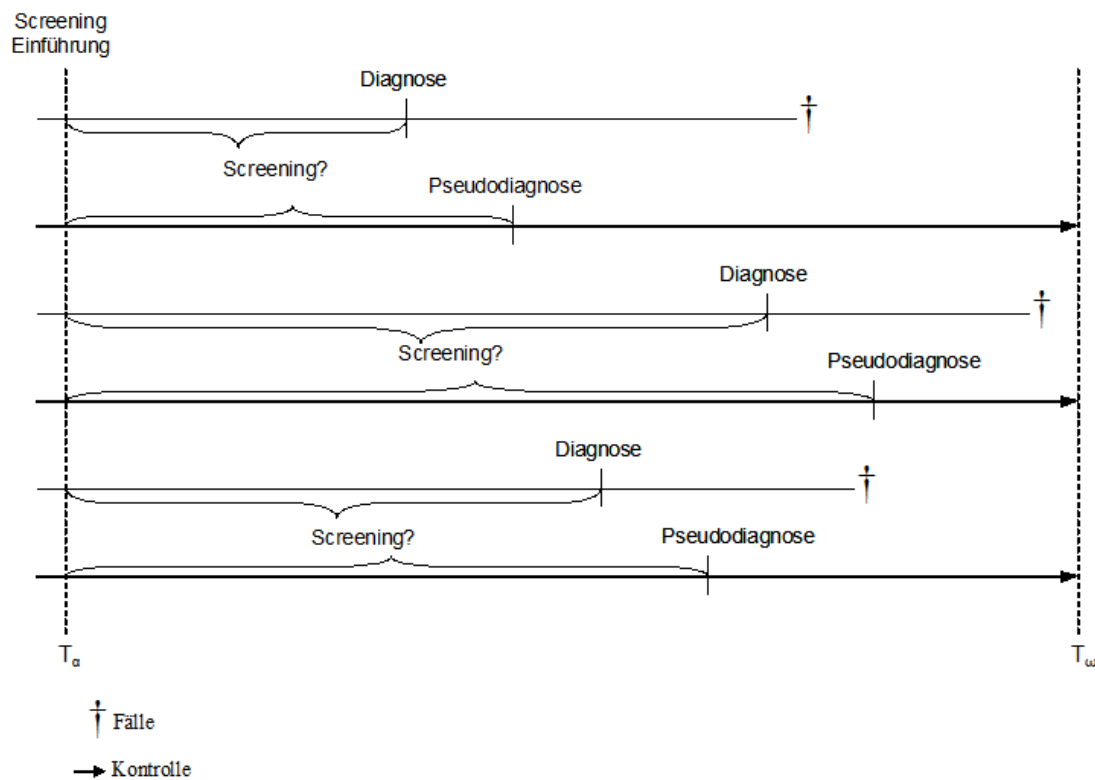


Abbildung 5. Fall-Kontroll-Studie mit Verschiebung des Pseudodiagnosedatums der Kontrollen (Quelle: eigene Darstellung).

In der Evaluationen des schwedischen Programms der Arbeitsgruppe von Jonsson (28,29,30,31) wurde für die Lead-Time korrigiert. Dabei wurde zunächst der Überlebensunterschied zwischen beiden Gruppen (Intervention/Kontrolle) ermittelt, was als Lead-Time definiert wurde. Dann wurde ermittelt, wie viele Brustkrebstodesfälle in dieser Lead-Time unmittelbar vor Ende des Screeningsalters diagnostiziert wurden (siehe Abbildung 6). Das RR wurde dann wie folgt korrigiert:

$$RR_{\text{kor}} = RR - (\text{Brustkrebstodesfälle in } \textit{lead} \text{ Periode} / \text{Brustkrebs-Gesamtodesfälle})$$

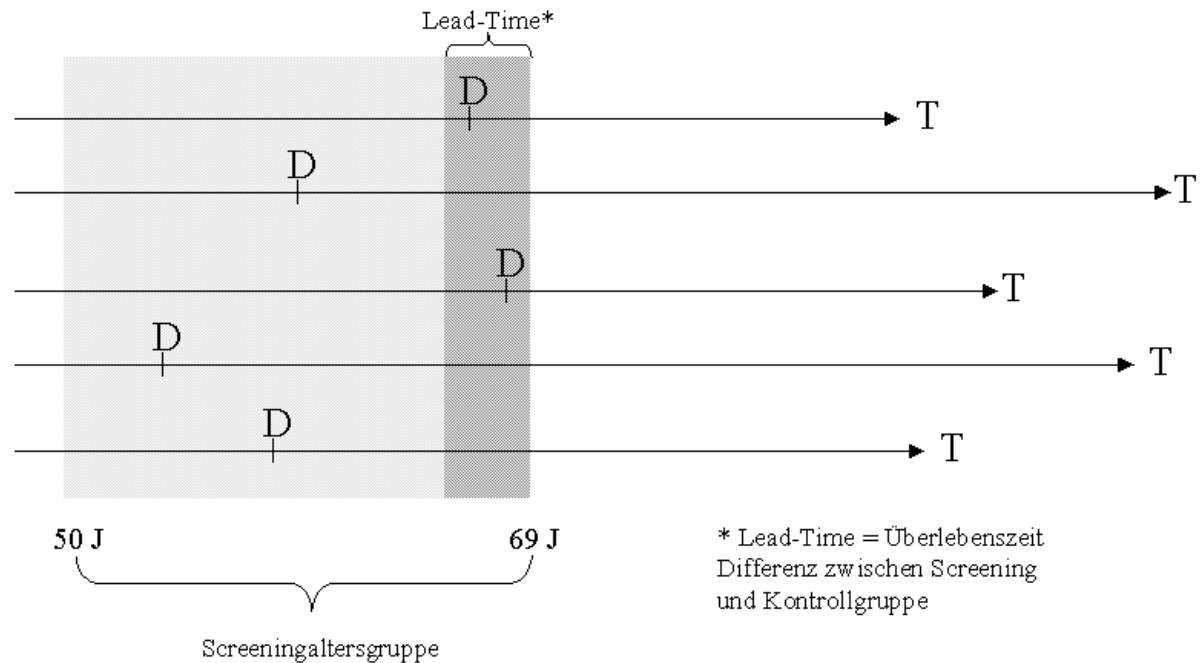


Abbildung 6. Umgang mit Lead-Time in Evaluationen des schwedischen Screenings (Bsp. für Screening Alter 50-69 J) (Quelle: eigene Darstellung).

In der Analyse der Effektivität des Screenings für die Altersgruppe 40-49 J. erwähnt die Studie von Jonsson, dass durch die Vorverlagerung der Diagnose (was auf 0,11 Jahre geschätzt wird) die Lead-Time sich zuungunsten des Screenings auswirken kann. Manche Frauen in den Screeningregionen (Regionen mit Screening 40-49 J.) hätten die Diagnose aufgrund der Mammographie kurz vor dem 50. Lebensjahr (konkret nach 49,89 Jahren) erhalten und ihr späterer Tod wurde damit der Interventionsgruppe zugeordnet. In den Kontrollregionen (kein Screening vor dem 50. Lebensjahr) wären diese Diagnosen erst nach dem 50. Lebensjahr gestellt worden, und ihr Tod nicht der Studiengruppe der Frauen zwischen 40-49 J. zugeordnet worden. Die Autoren schätzten, dass der Effekt des Screenings insgesamt dadurch um 0,4% unterschätzt worden wäre (30).

In der Analyse des Screenings-Programms in einer einzelnen schwedischen Region (Gävleborg) wurden die Ergebnisse für die Effekte von Lead-Time-Bias aufgrund der o.g. Überlegungen ebenfalls korrigiert. Dabei wurde zunächst der Überlebensunterschied zwischen beiden Gruppen ermittelt und dann das RR entsprechend korrigiert. Die Autoren errechneten eine Differenz von 1,07 Jahren in der Überlebenszeitdauer zugunsten der Screeninggruppe. Dann wurde ermittelt, wie viele der BK-Todesfälle in den 1,07 Jahren unmittelbar vor Beendigung des Screeningalters (also im Alter von 62,93 bis 64 Jahre) diagnostiziert worden waren. Im konkreten Beispiel der Gävleborg-Studie ergab diese Korrektur eine Reduktion des RR um 4% ($RR = 0,84$, ermittelte Lead-Time 1,07 Jahre, 10 BK-Todesfälle in Lead-Periode bei insgesamt 239 BK-Todesfälle in der Screeninggruppe, also 4,2% aller BK-Todesfälle ergibt RR_{kor} von 0,80) (31).

In der Studie derselben Arbeitsgruppe zur Altersgruppe 50-69 Jahren wurde dieselbe Methode verwendet, wobei ein Überlebensunterschied von 0,16 Jahren ermittelt wurde und die Autoren schlussfolgerten, dass das RR sich nur sehr geringfügig ändern würde (0,4%) (29). In der Evaluation für die Altersgruppe 70-74 J. wurde eine Lead-Time von 0,1 J. ermittelt und eine

entsprechende Reduktion des RR um 1,7% berechnet (32). In der Evaluation des Programms in Nordschweden wurde in ähnlicher Weise für Lead-Time adjustiert, ohne dass die Ergebnisse sich stark verändert hätten (Lead-Time 0,29 Jahre und RR um 2% reduziert), wobei auch die Möglichkeit eines Bias zugunsten von Screening berücksichtigt wurde (28).

In der Studie von Tabar (35,36), dass das schwedische Programm in den Provinzen Dalarna und Östergötland (Two-County-Studie) nach der flächendeckenden Einführung des Screenings analysiert, wurde angenommen, dass aufgrund des Lead-Time-Bias die Brustkrebsmortalität in der Interventionsgruppe unterschätzt und dadurch der Effekt des Screenings überschätzt hätte werden können. Diese Studie verglich die Brustkrebsmortalität in der Alterskohorte von 40-69 Jahren in der Periode vor Einführung des Screenings (1968-1977) mit der gleichen Alterskohorte im Zeitraum 1988-1996 (flächendeckende Einführung des Screenings mit einer Teilnahme von 85%). Im Zeitraum dazwischen wurde der Two-County-RCT durchgeführt, bei der 52% der Frauen randomisiert zum Screening eingeladen wurden. Während des RCT-Zeitraums wurden Brustkrebsfällen diagnostiziert, die, ohne Screening eigentlich im nächsten Zeitraum (also nach Ablauf des RCTs) diagnostiziert worden wären. Dementsprechend wären ohne vorhergegangenen RCT die Todesfälle dieser Erkrankungen in der Interventionsperiode berechnet worden. Unter der Annahme einer Lead-Time von 3 Jahren wurden die Todesfälle, deren Brustkrebs in den Jahren 1985-1987 diagnostiziert worden war, der Intervention zugerechnet. Dadurch wurde die potentielle Überschätzung durch Lead-Time in dieser speziellen Situation verringert (36). Insgesamt bedeutete dies in der Studie eine Reduktion des Screeningeffekts: RR 0,50 (0,41-0,60) vs. RR 0,58 (95%-CI 0,49-0,69) bei Berücksichtigung aller eingeladenen Frauen und RR 0,37 (95%-CI 0,30-0,46) vs. 0,44 (95%-CI 0,36-0,54), wenn nur die tatsächlich gescreenten Frauen berücksichtigt wurden (36). Das Vorgehen wird in Abbildung 7 dargestellt; um für Lead-Time zu korrigieren, wurden die Todesfälle T_{lead} der Interventionsgruppe zugeschrieben.

Tc = BK-Todesfälle Kontrollgruppe
 Ti = BK-Todesfälle Interventionsgruppe
 T0 = BK-Todesfälle nicht berücksichtigt
 Tlead = BK-Todesfälle berücksichtigt bei Interventionsgruppe um für Lead Time zu korrigieren

D = Zeitpunkt BK-Diagnose
 → = lebt

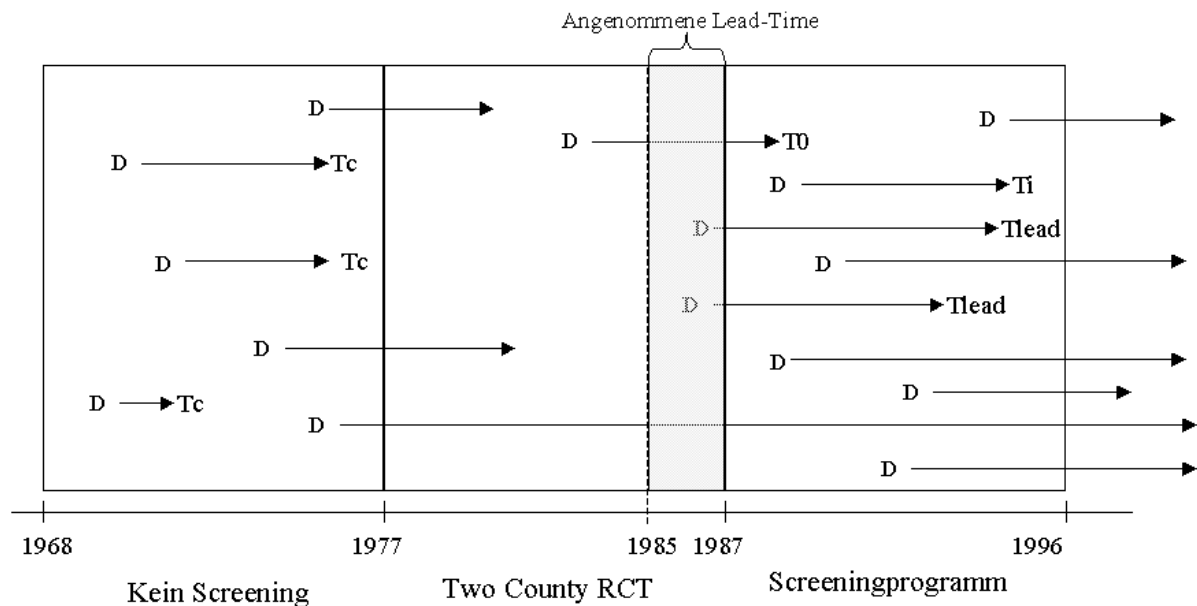


Abbildung 7. Umgang mit Lead-Time-Bias Studie von Tabar 2001 (Quelle: eigene Darstellung).

In den Studien, wo es vor der Screeningperiode möglich war, dass manche Fälle auch durch Screening entdeckt wurden (also in Situationen ähnlich zu der für die Two-Counties beschriebenen), kann alternativ die Anzahl der Personenjahre in der Interventionsgruppe reduziert werden, um den in der Lead-Time diagnostizierten Tumoren Rechnung zu tragen (27). Dadurch wird der Nenner für die Ermittlung der Mortalitätsrate reduziert, und diese entsprechend nach oben korrigiert. Dieser Ansatz wurde auch in einer schwedischen Studie verfolgt, bei der in manchen Regionen eine sehr große Kontamination (Anteil von Frauen die Screening Mammographien in der Vor-Screeningperiode erhalten hatten) vorlag. Die Korrektur verringerte den Effekt des Screenings in der betroffenen Region von RR 0,72 auf RR 0,84 (27).

3.4.2 Length-Time- / Over-Diagnosis-Bias

Nur wenige Studien verknüpfen in ihrer Beschreibung explizit spezifische Studiendesignmerkmale zu der Problematik von Length-Time- bzw. Over-Diagnosis-Bias (die verwandte Biasarten sind).

Die Verwendung der inzidenzbasierten Mortalität wird von manchen Autoren explizit als Strategie für die Minimierung von Length-Time- bzw. Over-Diagnosis-Bias bezeichnet (16,33,34). Die Evaluationsgruppe des schwedischen Screenings liefert die Begründung dafür wie folgt: durch Length-Time- bzw. Over-Diagnosis-Bias können die Anzahl von Brustkrebs-

diagnosen zwar erhöht werden, die Anzahl von Todesfälle (Zähler des Mortalitätsbruchs) wird jedoch dadurch nicht beeinflusst, genauso wenig wie die Personen-Jahren (Nenner des Mortalitätsbruchs) (33).

Die Verwendung von Exzess-Mortalität würde der höheren Anzahl an Brustkrebsdiagnosen aufgrund von Screening Rechnung tragen (29,30,31,32) und kann deshalb als eine Strategie zur Minimierung potentiellen Bias durch Überdiagnose beschrieben werden.

In der Überlebenszeit-Analyse des italienischen Screenings wurde für die Tumorcharakteristika adjustiert um ein potentiellen Length-Bias bzw. Over-Diagnosis-Bias zu berücksichtigen (15).

Thompson argumentiert, dass in Kohortenstudien Length-Time-Bias die Ergebnisse nicht beeinflussen kann, wenn die Definition der Kohorte nicht mit dem Diagnosezeitpunkt zusammenhängt und Mortalität untersucht wird (44). In ihrer Evaluation eines Screening-Programms in den USA verwendeten sie ein Fall-Kohorten-Studiendesign, was auch diesen Bias minimieren soll (44).

In der Evaluation des schwedischen Programms in der Two-County-Studie (35,36) wurde bei einer der Publikationen (35) eine Adjustierung vorgenommen, um Veränderungen in der Brustkrebsinzidenz zwischen der Prä-Screening und der Screeningperiode zu berücksichtigen. Dabei wurden die Ergebnisse für Veränderungen in der Brustkrebsinzidenz wie folgt adjustiert:

$$RR_{adj} = RR_m / RR_i$$

RR_{adj} bezeichnet das adjustierte RR für Mortalität, RR_m das nicht adjustierte Mortalitätsrisiko und RR_i das relative Risiko für Brustkrebsinzidenz. Letzteres wird aus dem Vergleich der Brustkrebsinzidenz in der Prä-Screeningkohorte und in der Screeningkohorte ermittelt. Diese Adjustierung war nach den Autoren notwendig, um der Erhöhung der Brustkrebsinzidenz, die auch unter den nicht-gescreenten Frauen zu beobachten war, Rechnung zu tragen (35).

Die Evaluationsgruppe des schwedischen Screenings entwickelte eine Methode, um alle Brustkrebstodesfälle zu berücksichtigen (unabhängig von Diagnosejahr – also nicht nur die Inzidenzbasierte Mortalität) und gleichzeitig für Veränderungen in der Brustkrebsinzidenz bzw. Letalität zu adjustieren. Die Autoren bildeten ein Regressionsmodell, in dem Variablen für die zeitlichen Trends in Inzidenz und Letalität, sowie ein Korrekturfaktor für Lead-Time eingingen (34). Dieses Modell lieferte konservativere Schätzer als wenn nur die Todesfälle eingeschlossen wurden, deren Brustkrebs nach der Einführung des Screening-Programms diagnostiziert worden war (siehe Tabelle 3).

Tabelle 3. Unterschied zwischen Inzidenzbasierter Mortalität und Todesfällen unabhängig von Diagnosezeitpunkt

Publikation	Mortalität	RR (95%-CI)	
		<i>self-selection</i> nicht korrigiert	<i>self-selection</i> korrigiert
SOSSEG 2006a ³³	nur Todesfälle mit Diagnose nach Einführung Screening	0,55 (0,51-0,59)	0,57 (0,53-0,62)
SOSSEG 2006b ³⁴	alle Todesfälle unabhängig von Diagnosedatum	0,58 (0,53-0,62)	0,61 (0,55-0,68)

Diese Ergebnisse suggerieren, dass der Ausschluss von Brustkrebstodesfällen, deren Diagnose vor der Einführung des Screenings gestellt wurde, und die Nicht-Berücksichtigung von Inzidenz- und Letalität-Trends zu einer gewissen Überschätzung der Screeningeffekte führen können.

3.4.3 Self-Selection-Bias

Frauen, die in Screening-Programmen für Brustkrebs teilnehmen, haben im Schnitt einen höheren sozioökonomischen Status als solche, die nicht teilnehmen. Ein höherer sozioökonomischer Status ist mit einem höheren Risiko für die Entwicklung von Brustkrebs verbunden. Allerdings ist das Risiko an Brustkrebs zu sterben geringer als bei Frauen mit einem niedrigeren sozioökonomischen Status (41). Insgesamt führt diese Konstellation zu einer Überschätzung des Effekts des Screening-Programms, weil Frauen mit einem geringeren Mortalitätsrisiko eher am Screening teilnehmen. In manchen Studien wird dieser Bias auch *healthy screen participation bias* oder *healthy volunteer bias* bezeichnet.

Das Problem des Self-Selection-Bias bei der Evaluation von Screening-Programmen kann dadurch reduziert werden, dass man als Interventionsgruppe die eingeladenen Frauen berücksichtigt, unabhängig davon, ob sie am Screening teilnahmen oder nicht. In insgesamt 21 Studien wurde die Exposition so gemessen. Wenn stattdessen als Interventionsgruppe die tatsächlich gescreenten Frauen berücksichtigt werden, werden die Effekte des Programmes aufgrund von Self-Selection überschätzt.

In der Evaluation des finnischen Programms (einer kontrollierten, quasi-randomisierten Studie) wurde die Gruppe der eingeladenen Frauen (unabhängig davon, ob sie tatsächlich gescreent wurden oder nicht) mit der Geburtskohorte, die nicht eingeladen wurde, verglichen. Die Autoren nennen diese Form der Analyse zutreffend „*intention-to-screen*“ (9). Die Autoren berichten zusätzlich die Brustkrebsmortalität derjenigen Frauen, die eingeladen wurden, aber nicht teilnahmen. Dabei zeigt sich, dass ein Vergleich zwischen tatsächlich gescreenten und nicht-gescreenten Frauen zu einer sehr großen Überschätzung der Effektivität des Programmes geführt hätte. So war die Brustkrebs-SMR (Standardisierte Mortalitätsrate) bei Frauen, die gescreent wurden, 0,27, während sie bei Frauen, die eingeladen wurden und nicht teilnahmen, doppelt so hoch lag (BK-SMR = 0,58). Noch stärker war der Unterschied, wenn auch Fälle, deren Diagnosezeitpunkt vor dem Zeitpunkt der Programmimplementierung lag, berücksichtigt wurden (0,63 respektive 3,67) (9,10) (siehe Tabelle 4). Diese Ergebnisse zeigen, wie stark durch Self-Selection die Ergebnisse verzerrt werden können, wenn keine *Intention-to-screen* Analyse durchgeführt wird.

Tabelle 4. Intention-to-screen (ITS) Analyse vs. „per Protokoll“ (pP) Analyse (tatsächlich gescreente).

Studie	SMR (alle BK-Todesfälle)						SMR (nur BK-Todesfälle BK-Diagnose während Program)					
	Eingeladene			Kon- trolle	RR (ITS)	RR (pP)	Eingeladene			Kon- trolle	RR (ITS)	RR (pP)
	Gescreent	Nicht Gescreent	Total				Gescreent	Nicht Gescreent	Total			
Hakama ^{9,10}	0,63	3,67	1,01	1,12	0,90	0,56	0,27	0,58	0,31	0,41	0,76	0,66

Quelle: eigene Darstellung mit Angaben aus den Publikationen

In Studien, in denen die Exposition durch die tatsächliche Teilnahme am Screening definiert wird, wie dies bei den vorliegenden Fall-Kontroll-Studien häufig war, wurde eine Korrektur der Ergebnisse durchgeführt, um das Problem der Self-Selection zu berücksichtigen.

Eine Methode, um die Ergebnisse zu korrigieren, ist von Duffy vorgeschlagen (48) und in mehreren Studien angewandt worden. Dabei werden die Teilnahmeraten (beobachtete oder angenommene) und das erhöhte Mortalitätsrisiko der zum Screening eingeladenen aber nicht teilnehmenden Frauen (was z.B. aus randomisierten Studien ermittelt werden kann) berücksichtigt. Die von Duffy et al. vorgeschlagene Korrektur folgt der folgenden Formel:

$$OR_{\text{korr}} = T \times OR \times D_{nT} / (1 - (1 - T) D_{nT})$$

wobei T die Teilnahmerate und D_{nT} das (erhöhte) Sterblichkeitsrisiko in der Gruppe der Nicht-Teilnehmerinnen gegenüber der nicht zum Screening eingeladenen darstellen. OR bezeichnet die nicht korrigierte OR.

Als Ergebnis der Korrektur steht eine korrigierte OR, die um einen Faktor f größer als die rohe OR ist. Durch die Korrektur wird also eine konservativere Schätzung des Effekts vorgenommen. Tabelle 5 listet die verwendeten Werte der Parameter und den jeweils resultierenden Korrekturfaktor in den Studien auf, die diese Methode verwendet haben (4,12,17,27,40,41). In der Tabelle wird ersichtlich, dass in manchen Studien die statistische Signifikanz durch die Korrektur erlischt, da das Konfidenzintervall dann 1 mit einschließt.

Tabelle 5. Korrektur für Self-Selection-Bias (Beispiele).

Studie	Teilnahmerate	Mortalitätsrisiko Nicht-Teilnehmerinnen (D_{nT})	“Korrekturfaktor” (f)	nichtadjustierte OR (95%-CI)	Für self-selection adjustierte OR (95%-CI)
Coldman 2006 ⁴	71%	17%	1,26	0,60 (0,55-0,65)	0,76 (k.A.)
Fielder 2004 ⁴¹	77%	36%	1,53	0,49 (0,36-0,66)	0,75 (0,49-1,14)
Gabe 2007 ¹²	68%	17%	1,27	0,59 (0,41-0,84)	0,75 (0,52-1,09)
Allgood 2008 ⁴⁰	k.A.	k.A.	1,86	0,35 (0,24-0,50)	0,65 (0,48-0,88)
Puliti 2008 ¹⁷	65%	11%	1,19	0,46 (0,38-0,56)	0,55 (0,36-0,85)
Roder 2008 ²	k.A.	17%	1,19	0,59 (0,47-0,74)	0,70 (k.A.)
Tabar 2001 ³⁶	85%	34%	1,40	0,37 (0,30-0,46)	0,52 (0,43-0,63)
SOSSEG ³⁴	78%	17%	1,05	0,58 (0,53-0,62)	0,61(0,55-0,68)

Quelle: eigene Darstellung mit Angaben aus den Publikationen

In den Arbeiten von Tabar wurde für die Ergebnisse einer Kohortenstudie eine ähnliche Korrektur vorgenommen, wobei das RR wie folgt korrigiert wurde (35,36):

$$RR_{\text{korr}} = T \times RR + (1 - T) D_{nT}$$

In dieser Korrektur stellt T die Teilnahmerate am Programm und D_{nT} das Sterblichkeitsrisiko bei Nicht-Teilnehmerinnen im Vergleich zu nichteingeladenen Frauen (ermittelt aus einem RCT) dar.

Die Evaluation der schwedischen Screening Evaluation Group nimmt eine Korrektur für Self-Selection vor, die sich an der o.g. Duffy-Methode anlehnt, aber diese modifiziert. Die

Modifizierung wird dadurch begründet, dass in der Studie bei den Nicht-Teilnehmerinnen sich meistens um nicht eingeladene Frauen handelte:

$$RR_{\text{kor}} = T \times RR / (1 - (1 - T) D_{\text{nT}})$$

wobei T die Teilnahmerate am Programm und D_{nT} das Sterblichkeitsrisiko bei Nicht-Teilnehmerinnen im Vergleich zu nicht-eingeladenen Frauen darstellt. In diesem Fall wurde die D_{nT} nicht basierend auf Daten einer RCT angenommen, sondern aus der Studie selbst, in dem das RR für die nicht exponierten Frauen während der Screeningphase verwendet wurde (33,34).

Die Definition der Exposition als „gescreent“ (= Kontrolle „nicht gescreent“) tendiert dazu, stärkere Effekte zu zeigen. Diese Überschätzung wird durch die o.g. Korrekturen zwar reduziert, allerdings liefert die Definition der Exposition als „eingeladen“ (= Kontrolle „nicht eingeladen“) die konservativere Schätzungen des Effektes von Screening, wie aus der Ergebnissen der Studien, die diese Sensitivitätsanalysen durchführten, hervorgeht (17,27,40,42) (siehe Tabelle 6).

Tabelle 6. Self-Selection – Definition der Exposition und Korrektur

Studie	RR/OR (95%-CI)		
	„tatsächlich gescreent vs. nicht gescreent“	„gescreent vs. nicht gescreent“ korrigiert für self-selection	„eingeladen vs. nicht eingeladen“
Puliti 2008 ¹⁷	0,46 (0,38-0,56)	0,55 (0,36-0,85)	0,75 (0,62-0,92)
Duffy 2002 ²⁷	0,56 (0,50-0,62)	0,61 (0,55-0,68)	0,68 (0,60-0,77)
Moss 1992 ⁴²	0,51 (0,27-0,98)	-	0,76 (0,54-1,08)
Allgood 2005 ⁴⁰	0,35 (0,24-0,50)	0,52 (0,32-0,84)	0,65 (0,48-0,88)
Tabar 2001 ³⁶	0,37 (0,30-0,46)	0,52 (0,43-0,63)	0,50 (0,41-0,60)
SOSSEG ^{33;34}	0,55 (0,51-0,59)	0,57 (0,53-0,62)	0,73 (0,69-0,77)

Quelle: eigene Darstellung mit Angaben aus den Publikationen

In den Studien zum dänischen Mammographie-Screening-Programm führten die Autoren eine Subgruppenanalyse für diejenige Population durch, die am Screening-Programm teilgenommen hatte (5,6). Dabei wurde das relative Risiko für Teilnehmenden am Screening für Self-Selection-Bias korrigiert, indem es durch einen Korrekturfaktor dividiert wurde, der über die folgende Gleichung berechnet wurde:

$$1 = RR_{\text{corrected}} * w_{\text{participants}} + RR_{\text{non-attenders}} * w_{\text{non-attenders}}$$

$w_{\text{participants}}$ entspricht dem Anteil der Gesamtanzahl von Personenzahlen im Screeningzeitraum, den die Teilnehmenden beigetragen und $w_{\text{non-attenders}}$ ist der Anteil an Personenzahlen der Nicht-Teilnehmenden. $RR_{\text{non-attenders}}$ entspricht dem relativen Risiko für die Nicht-Teilnehmenden, das sich genauso berechnet wie für die eingeladene Population, aber nur die Nicht-Teilnehmenden einbezieht. Das korrigierte relative Risiko für die Teilnehmenden spiegelt den Effekt der Screeningteilnahme für die Population der Teilnehmenden wider (5,6).

In der Studie von Moss (42) wurde keine Korrektur bzw. Adjustierung für Self-Selection-Bias vorgenommen. Es wurde aber versucht das Ausmaß einer potentiellen Verzerrung durch dieses Problem abzuschätzen. Dafür wurden die Ergebnisse von zwei Analysen berichtet: 1) Fall-Kontroll-Analyse mit Fällen und Kontrollen aus einer Region mit und aus einer Region ohne Screening-Programm, 2) Fall-Kontroll-Analyse mit Fällen und Kontrollen nur aus der Region mit

Screening-Programm. Dies bedeutet, dass in der 1. Analyse sowohl eingeladene (Region mit Screening) als auch nicht eingeladene Frauen (Region ohne Screening) berücksichtigt wurden. In der 2. Analyse wurden alle Frauen eingeladen. Der Vergleich des Brustkrebs-Mortalitätsrisikos in der Region ohne Screening-Programm mit dem Brustkrebs-Mortalitätsrisiko der nichtgescreenten Frauen (d.h. der Nicht-Teilnehmerinnen) in der Region mit Screening-Programm weist auf die Möglichkeit von Self-Selection-Bias hin. Die Nicht-Teilnehmerinnen in der Region mit Screening hatten ein höheres Brustkrebs-Mortalitätsrisiko als Frauen in der Region ohne Screening-Programm (die dementsprechend auch nicht gescreent wurden) (OR 1,17 [95%-CI 0,71-1,93]) (42). Die Studie zeigt, dass ein Vergleich zwischen gescreenten und nichtgescreenten Frauen innerhalb einer Gruppe von eingeladenen Frauen für Self-Selection-Bias anfällig ist. Die Ergebnisse der ersten Analyse sind daher konservativer als die der zweiten (OR 0,76 [95%-CI 0,54-1,08] vs. OR 0,51 [95%-CI 0,27-0,98]) (42).

3.4.4 Verzerrungen im Zusammenhang mit dem Diagnosezeitpunkt und der Exposition

In den Studien können Verzerrungen entstehen, wenn der Diagnosezeitpunkt der Brustkrebsfälle den Zusammenhang zwischen Exposition und Mortalitätsrisiko beeinflussen kann. In den Studien sind unter diesen Gesichtspunkt verschiedene, jedoch verwandte, Biasmöglichkeiten beschrieben worden. Die Strategien, um die folgenden Verzerrungsquellen zu berücksichtigen, sind entsprechend verwandt und teilweise überlappend. Die Bezeichnungen der Bias in den verschiedenen Studien sind dementsprechend teilweise untereinander überlappend bzw. austauschbar.

Time of Diagnosis Bias

Lange Überlebenszeiten sind bei Brustkrebs nicht selten, so dass viele Sterbefälle während der Beobachtungszeit auf Brustkrebsfälle zurückzuführen sind, die vor der Implementierung des Screenings-Programms diagnostiziert wurden. Da Frauen mit einer Brustkrebsdiagnose nicht zum Screening eingeladen werden, kann diese Situation in Fall-Kontroll-Studien zu einer Überschätzung der Effekte des Screenings führen, da diese in der Studienzeit beobachteten Todesfälle in Verbindung mit der Nichtteilnahme am Screening gebracht werden würden (Abbildung 8).

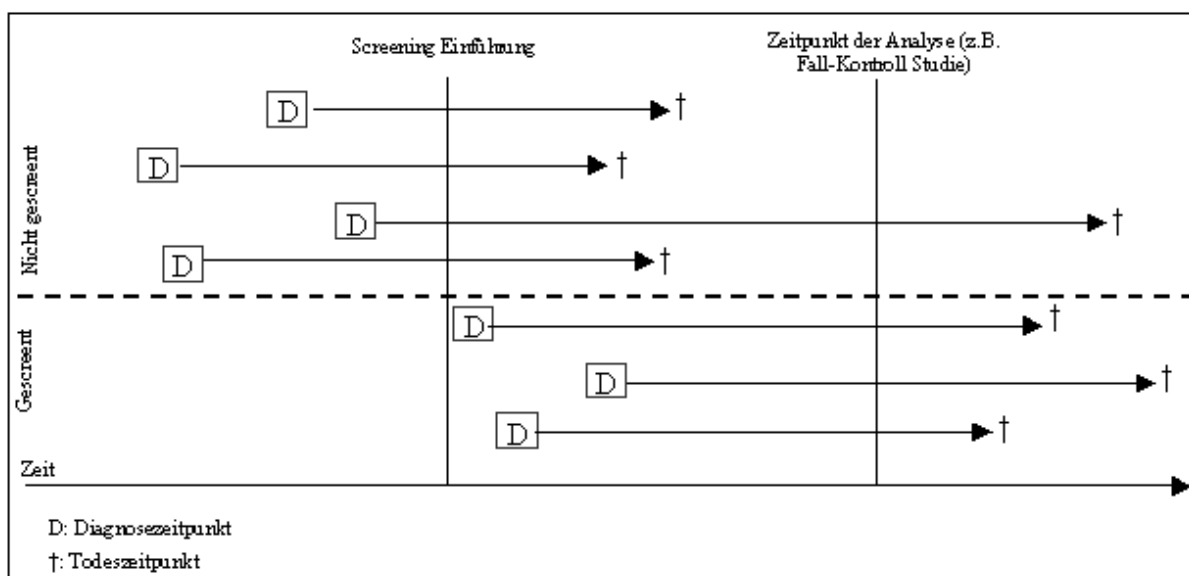


Abbildung 8. Time of Diagnosis Bias in Fall-Kontroll-Studien (Quelle: eigene Darstellung).

Dieser Bias wird in der Studie von Fielder 2004 *time of diagnosis bias* genannt, wobei – um ihn zu vermeiden – die Autoren alle Frauen ausgeschlossen haben, deren Brustkrebsdiagnose vor dem Zeitpunkt der Programmimplementierung lag (41). Der Ausschluss von Brustkrebsdiagnosen, die vor der Einführung des Screening-Programms gestellt wurden, wird auch in weiteren Studien berichtet (siehe Tabelle 12, Spalte „Andere“), ohne dabei den adressierten Bias explizit *time of diagnosis bias* zu nennen. Andere Autoren beschreiben diesen Ausschluss als die Anwendung von *incidence based mortality* oder *refined mortality*, wobei damit auch potentiellen Bias durch Überdiagnose minimiert werden soll (siehe Abschnitt 3.3). Im Zusammenhang mit der Tatsache, dass ein Teil der Sterbefälle während der Beobachtungszeit auf Brustkrebsfälle zurückzuführen sind, die vor der Implementierung des Screening-Programms diagnostiziert wurden, wird auch die Problematik diskutiert, dass ein Screening-Programm seinen vollständigen potentiellen Effekt erst einige Jahre nach Implementierung entfaltet. Wenn die Analyse alle Brustkrebstodesfälle einschließt (also auch solche, die auf Diagnosen vor der Implementierung oder in der Anlaufphase des Programms zurückzuführen sind), kann der Effekt des Screenings unterschätzt werden, denn es werden Todesfälle eingeschlossen, bei denen das Programm keinen Effekt haben konnte (9).

In der Evaluation des finnischen Programms wurden Analysen mit und ohne Berücksichtigung der Todesfälle durchgeführt, bei denen die Brustkrebsdiagnose vor dem Zeitpunkt der Einführung des Screenings gestellt worden war. Bei diesen Studien mit quasi-randomisiertem, kontrollierten Design zeigt sich, dass der Einschluss aller Brustkrebstodesfälle im Beobachtungszeitraum zu schwächeren Effekten führte, als wenn nur die Brustkrebstodesfälle, deren Brustkrebsdiagnose nach Implementierung des Programms gestellt worden war, berücksichtigt wurden (9,10,11) (siehe Tabelle 7).

Tabelle 7. Vergleich RR mit und ohne Berücksichtigung der Todesfälle mit BK-Diagnose vor Einführung des Programms.

Studie	RR (95%-KI) Alle BK-Todesfälle	RR (95 %-KI) Nur BK-Todesfälle, die nach dem Zeitpunkt der Screeningeinführung diagnostiziert wurden
Hakama ¹¹	0,27 (0,07-0,81)	0,11 (0,00-0,71)
Hakama ^{9,10}	0,90 (k.A.)	0,76 (0,53-1,09)

Quelle: eigene Darstellung aus Angaben in den Publikationen

Setup-Bias/Inclusion-Bias

Am Anfang der Programme werden aus organisatorischen und logistischen Gründen noch nicht alle Frauen der Zielgruppe eingeladen. Es dauert eine Zeit, bis die erste Screeningrunde vollständig eingeladen wurde. In dieser Zeit jedoch werden schon Brustkrebsdiagnosen erstellt (weil sie symptomatisch werden). Wenn die Zuordnung zu Screeninggruppe nicht anhand der tatsächlichen Einladung basiert, sondern aufgrund von Alter oder Wohnort unter der Annahme, dass eine Einladung erfolgt sein sollte, würden auch solche Fälle derr Screeninggruppe zugerechnet, die aufgrund von Schwierigkeiten, die mit der Beginn der Implementierung der Programme zusammenhängen, noch nicht eingeladen wurden. Dadurch kann eine Unterschätzung des Screeningeffektes entstehen. Je nach Autoren, wird dieses Problem Setup- (41) oder Inclusion-Bias bezeichnet (29,30). Um Einfluss auf die Ergebnisse durch solche Probleme zu reduzieren, können die Fälle, die in einer bestimmten Zeit nach dem Start der Implementierung

des Programms diagnostiziert wurden, ausgeschlossen werden. Bei dieser Strategie werden alle Brustkrebsdiagnosen aus der Anfangszeit ausgeschlossen (unabhängig davon ob sie tatsächlich schon eingeladen wurden oder nicht). In der Studie von Coldman (4) wurden alle Frauen ausgeschlossen, deren Diagnose in den ersten 6 Monaten nach Implementierung des Programms gestellt wurde. In der Studie von Fielder (41) wurden die ersten 3 Jahre ausgeschlossen.

In einigen der Evaluierungen aus Schweden werden alle Brustkrebsdiagnosen, die nach dem Start der Implementierung diagnostiziert wurden, eingeschlossen, aber eine Korrektur der Mortalitätsergebnisse vorgenommen, um den potentiellen Bias durch die noch nicht eingeladenen Frauen zu reduzieren. Für die Korrektur werden die erwarteten BK-Todesfällen unter den Frauen, die erst nach der Diagnose ihre Einladung erhielten, abgeschätzt (z.B. mittels Simulation unter Berücksichtigung der benötigte Zeit um die erste Screeningrunde zu beenden)(29,30,31,32). Dann wird der Effekt wie folgt korrigiert:

$$RR_{\text{kor}} = (O - \Psi) / [(O/RR) - \Psi]$$

wobei O die beobachteten Todesfällen während Studienzeitraum, Ψ die erwarteten Todesfällen bei Frauen die diagnostiziert wurden bevor sie eine Einladung erhielten und RR das nicht korrigierte relatives Risiko darstellen. Die Korrektur für diesen Bias wirkt sich wie in Tabelle 8 dargestellt aus.

Tabelle 8. Korrektur für Setup-/Inclusion-Bias (Beispiele).

Studie	RR	O	Ψ	RR _{kor}
Jonsson 2000 ³⁰	0,91	235	56	0,88
Jonsson 2001 ²⁹	0,84	648	118,1	0,81
Jonsson 2003a ³¹	0,82	-	4,4	k.A. ^a
Jonsson 2003b ³²	0,82	311	76,8	0,77

a Die Autoren berichten, dass aufgrund des niedrigen Anteils der vor Einladung diagnostizierten Brustkrebs-Fälle das Ergebnis durch diesen Bias kaum beeinflusst werden würde.
Quelle: eigene Darstellung nach Angaben in den Studien

In der ökologischen Studie von Törnberg, die den Effekt des Mammographie-Screening-Programms in Schweden untersuchte, wurde ein sogenannter Mammographie-Programm-Extension-Score gebildet (37). Dieser Score soll den Grad an Implementierung, der in jeder Provinz erreicht wurde, abbilden. Dabei wurde angenommen, dass zwischen 5 und 10 Jahren nach Start des Programms dieses 50% seines potentiellen Effekts erreicht. Erst ab 10 Jahre wird den vollen Effekt erreicht (100%), während zwischen 0 und 5 Jahre angenommen wurde, dass das Programm keinen Effekt haben konnte. Der Score bildete sich aus diesem Faktor mal dem Anteil an der Frauenbevölkerung, der eingeladen wurde (z.B. in einer Provinz waren nach 5 Jahren 16% der Frauen eingeladen, was einen Mammographie-Programm-Extension-Score von 0,08 ergab (16% x 50%)).

Exposure Opportunity Bias/Misclassification of Exposure Bias/Screen Detected Bias

Dieser Bias entsteht in Fall-Kontroll-Studien, weil ein Teil der Fälle (die in Fall-Kontroll-Studien per definitionem gestorben sind) durch das Screening entdeckt werden. Wenn dieses diagnostische Screening in der Screeninggeschichte berücksichtigt wird, wird die Exposition zu Screening der Fälle überschätzt und dementsprechend der Nutzen des Screenings unterschätzt (49). Dieses

Problem entsteht, weil die Fälle sowohl die Möglichkeit haben, in der Zeit vor Diagnose gescreent zu werden als auch zusätzlich die Möglichkeit im Diagnoseprozess gescreent zu werden (12). In den meisten Studien ist nicht möglich diagnostische Mammographien von echten Screeningmammographien zu unterscheiden (46). Bei den Kontrollen werden jedoch keine diagnostischen Screenings vorliegen, weil sie meistens frei von Brustkrebs sind (siehe Tabelle 2). Der Einschluss des Screenings, bei dem bei den Fällen die Erkrankung diagnostiziert wurde, führt zur Unterschätzung des Effektes des Screenings. Der Ausschluss dieses Detektions-Screenings (d.h. die Fälle, die nur ein Detektions-Screening haben, werden als „nicht gescreent“ klassifiziert), würde allerdings zu einer Überschätzung der Screening-Effekte führen (12, 41). In der Studie von Gabe 2007 aus Island (12) wurde der Effekt von Screening unter Berücksichtigung und unter Ausschluss des „diagnostischen Screenings“ berechnet, um einen „optimistischen“ und einen „konservativen“ Effektschätzer zu ermitteln (Tabelle 9). Dies wurde auch in der Evaluation des niederländischen DOM-Screening-Programms gemacht (in dieser Studie wird dieser Bias als *misclassification of exposure* bezeichnet). Nach dem Vergleich beider Ergebnisse schlussfolgerten die Autoren, dass der tatsächliche Effekt zwischen beiden Extremen liegen würde (24).

Tabelle 9. Exposure Opportunity Bias (Beispiele).

Studie	„Diagnostik-Screen“ als Exposition OR (95%-CI)	„Diagnostik-Screen“ nicht als Exposition OR (95%-CI)
DOM-Projekt ²⁴	0,54 (0,37-0,74)	0,38 (0,26-0,56)
Gabe 2007 ¹²	0,59 (0,41-0,84)	0,34 (0,24-0,49)

Eine weitere Möglichkeit ist die Korrektur der Ergebnisse für diesen Bias nach einer von Duffy et al. (49) vorgeschlagene Methode. Dabei wird ein Korrekturfaktor aus dem Screeningintervall, die Screeningteilnahme in der Gruppe der Fälle, und die Proportion der „Prävalenz-Screens/Detektion-Screens“ zu allen Screenings nach folgender Formel ermittelt:

$$c^* = \frac{\lambda x}{(p_1 Se^{-\lambda x} + \lambda x)}$$

und dieser geht als Korrekturfaktor in folgender Gleichung ein:

$$\ln(\text{OR}_{\text{kor}}) = \ln(\text{OR}) + \ln(c^*) - (1 - c^*)p$$

wobei p der Anteil von Fällen die als gescreent klassifiziert wurden (einschließlich Detektions-Screening) darstellt (49). Die in dieser Form korrigierte OR war 0,51 (95%-CI 0,31-0,85) (12) und lag damit zwischen den optimistischen und den konservativen Schätzer (siehe Tabelle 9).

In der Studie von Allgood (40) soll nach den Autoren der Ausschluss von Fällen, die in den ersten Jahren nach der Implementierung des Programmes diagnostiziert wurden, diesen Bias (in dieser Studie *Screening Opportunity Bias* genannt) minimieren. Es wurden nur Fälle (d.h. an Brustkrebs verstorbene Frauen) eingeschlossen, deren Brustkrebsdiagnose frühestens 1995 gemacht wurde (d.h. frühestens 6 Jahre nach der Implementierung des Programmes). Diese Strategie beruht auf der Annahme, dass in den ersten Jahren nach Beginn eines Screening-Programms (die Prävalenzrunde) die Anzahl von „screen-detected“ Fällen höher ist (40,41). Andere Autoren beschreiben diese Strategie (Einschluss von Fällen erst ab einem bestimmten Zeitpunkt nach der Einführung des Screenings) als geeignet, um den Effekten von Lead-Time-Bias zu begegnen (24).

In diesen Zusammenhang ist auch relevant, in wie weit durch die Brustkrebsdiagnose bei Kontrollen die Exposition beeinflusst wird. In Fall-Kontroll-Studien könnte eine Verzerrung zuungunsten der Screeninggruppe entstehen, wenn in den Kontrollen eine Diagnose von Brustkrebs vorliegt, die vor dem Diagnosezeitpunkt der Fälle liegt. Das Vorliegen einer Brustkrebsdiagnose führt zum Ausschluss vom Screening, d.h. die Kontrollen würden kein Screening aufweisen, während die Fälle eins aufwiesen. Diese Situation würde die Ergebnisse zu ungunsten von Screening verzerren, da nicht Screening in Zusammenhang mit Überleben gebracht würde. Um diesen Bias zu vermeiden, sollen in Fall-Kontroll Studien Kontrolle ausgeschlossen werden, wenn sie eine Brustkrebsdiagnose aufweisen, die vor der Zeitpunkt der Diagnose des respektiven Falles liegt (2). Das Fehlen einer Brustkrebsdiagnose bei Kontrollen zum Zeitpunkt der Brustkrebsdiagnose der respektiven Fällen macht die Festlegung einer Pseudodiagnose um den Expositionszeitraum zu begrenzen notwendig (siehe Abschnitt 0).

3.4.5 Weitere Aspekte

Wie in Tabelle 13 dargestellt, werden in den Studien weitere Aspekte benannt, die bei der Evaluation von Screening-Programmen berücksichtigt werden müssen.

Beobachtete Unterschiede im Brustkrebs-Mortalitätsrisiko zwischen Kohorten können z.B. durch Unterschiede in der Wirksamkeit der Therapie oder in der Verteilung anderer Risikofaktoren. Diese Unterschiede können sowohl in Studien bei denen die Exposition in Zusammenhang mit der Zeit steht als auch in zeitlich parallelen Kohorten vorhanden sein. Wenn die Kohorten über die Zeit gebildet werden (Nicht-Exponierte vor und Exponierte nach Einführung des Screenings), dann können Unterschiede in der Mortalität z.B. durch die Verbesserung der Therapie über die Zeit zum Teil erklärt werden. Bei geographischen Vergleichen können z.B. regionale Unterschiede in der Qualität der Behandlung von Brustkrebs Unterschiede in der Mortalität zum Teil erklären.

Um das Ausmaß geographisch oder zeitlich bedingter potentieller Unterschiede in der Brustkrebsmortalität abzuschätzen, wird in einigen Kohorten-Studien die Brustkrebsmortalität in der Zeit vor Einführung des Screenings sowohl in der exponierten Kohorte als auch in der nicht exponierten Kohorte untersucht (siehe Seite 15). Die Ergebnisse in der Studienperiode können anhand der Ergebnisse in der sogenannten Referenzperiode dann adjustiert werden (siehe z.B. 28,31). Manche Studien vergleichen die Reduktion der Brustkrebsmortalität über die Zeit bei zum Screening exponierten Frauen und bei nicht exponierten (z.B. 35,38). Der Unterschied wird in diesen Studien dem Screeningeffekt zugeschrieben.

Andere Studien haben in multivariaten Analysen z.B. für Beobachtungszeitraum, Jahr der Diagnose, Jahr des Todes, Ort oder Durchschnittseinkommen im Wohnort adjustiert, um für den Einfluss von potentiellen zeitlichen bzw. geographische Unterschieden zu kontrollieren (z.B. 5,6,14,24,30,35,36).

Das Alter ist eine der wichtigsten Faktoren, um Mortalitätsunterschiede zu erklären, und wurde entsprechend in den meisten Studien berücksichtigt. Je nach Studiendesign und Datenverfügbarkeit sind in den vorliegenden Studien verschiedene Strategien dafür verwendet worden. In den Fall-Kontroll-Studien wurden die Fälle und Kontrolle nach Alter gematcht. In manchen Studien

ging das Alter in multivariate Modelle ein (z.B. 4,44) oder es wurde die altersstandardisierte Mortalität berechnet (z.B. 13,14,43).

Ein weiteres Problem, das manche Studien (z.B. 3,29,30) diskutiert haben, ist die potentialle Unterschätzung der Effekte von Screening-Programme beim Vorliegen von Kontamination. Kontamination entsteht, wenn in der Kohorte der Nicht-Exponierten bei einem Teil der Frauen Screening-Mammographien durchgeführt wurden (z.B. als opportunistisches Screening beim Gynäkologenbesuch), so dass eine gewisse Exposition doch stattgefunden hat. Diese Frauen zählen jedoch zur Kohorte der Nicht-Exponierten aufgrund ihrer Wohnorts oder der Zeitperiode. Vorausgesetzt, dass Screening positive Effekte hat, führt Kontamination dazu, dass die beobachteten Unterschiede geringer werden, als wenn in der Gruppe der Nicht-Exponierten tatsächlich keine Screening stattgefunden hätte.

In der dänischen Studie sprechen die Autoren von einem Bias durch *different follow-up periods* (5,6). Nach diesen Autoren entsteht dieser Bias dadurch, dass, wenn die prävalente Brustkrebsdiagnosen (bei Frauen, deren Brustkrebsdiagnose vor der Implementierung des Screenings gestellt wurde) ausgeschlossen werden, die Brustkrebsmortalitätsrate einer Studienpopulation in einem bestimmten Zeitraum nicht konstant bleiben wird. In den ersten Beobachtungsjahren werden weniger brustkrebsbedingte Todesfälle auftreten als in den darauffolgenden, da die prävalenten Fälle ausgeschlossen wurden. Olsen et al. vermeiden diesen Verzerrungsfaktor, indem Frauen in der gescreenten Gruppe und in den Kontrollgruppen über gleichlange Zeiträume beobachtet werden und bei den nicht exponierten Frauen ein Pseudo-Einladungssystem mit gleichen Einladungszeitpunkten wie in der exponierten gebildet wird (6). In den meisten Studien, die Kohorten von exponierten und nicht exponierten Frauen verglichen, wurde darauf geachtet, dass die Beobachtungszeiträume in beiden Kohorten ähnlich lang waren, ohne diese Maßnahme explizit als Strategie gegen einem spezifischen Bias zu bezeichnen.

3.5 Ergebnisse der Studien

Ziel dieser Arbeit ist nicht eine Bewertung der Effektivität der Mammographie anhand der vorliegenden Ergebnisse. Um einen vollständigen Überblick der Studien zu gewährleisten wurden jedoch auch deren Ergebnissen extrahiert. In der „Tabelle 14. Ergebnisse Brustkrebsmortalität“ werden die Hauptergebnisse der Studien zusammengefasst (Ergebnisse zu Brustkrebsmortalität). Es handelt sich um die Ergebnissen der Vergleichen zwischen Interventionsgruppen (eingeladen bzw. tatsächlich gescreent) und Kontrollgruppen (nicht eingeladen bzw. nicht gescreent). In einigen Studien wurden auch Ergebnisse nach der Anzahl der durchlaufenen Screeningrunden, des Alters beim ersten Screening oder dem Zeitpunkt des letzten Screenings (z.B. „weniger als 2 Jahren vs. mehr als 2 Jahren“) berichtet, die jedoch hier nicht extrahiert wurden.

Andere Studien berichteten Surrogatparameter, die zur Bewertung der Qualität der Screening-Programme (z.B. Tumorgröße, Histologie, Grading, Staging, Notwendigkeit von adjuvanter systemischer Therapie) herangezogen werden, da sie als Prädiktoren für Brustkrebsmortalität gelten (3).

Die Betrachtung der Ergebnisse dient auch dazu, den Einfluss verschiedener Strategien gegen Bias auf die Effektschätzer abzuschätzen. Wie zu erwarten, sind stärkere Mortalitätsreduktionen

zu beobachten, wenn tatsächlich gescreent vs. nicht gescreent verglichen wird, wobei diese dann durch die Korrekturen für Self-Selection abgeschwächt werden (siehe Abbildung 9). Die Effekte sind schwächer, wenn die Gruppe der eingeladenen mit der Gruppe der nicht eingeladenen Frauen verglichen wird. Letzterer Vergleich stellt die Effektivität der Screening-Programme dar.

In wie weit die verschiedenen Korrektur- bzw. Adjustierungsansätze für spezifischen Biasarten auf die Ergebnisse einzelner Studien sich auswirken können, ist in den jeweiligen Bias-Kapiteln exemplarisch beschrieben worden.

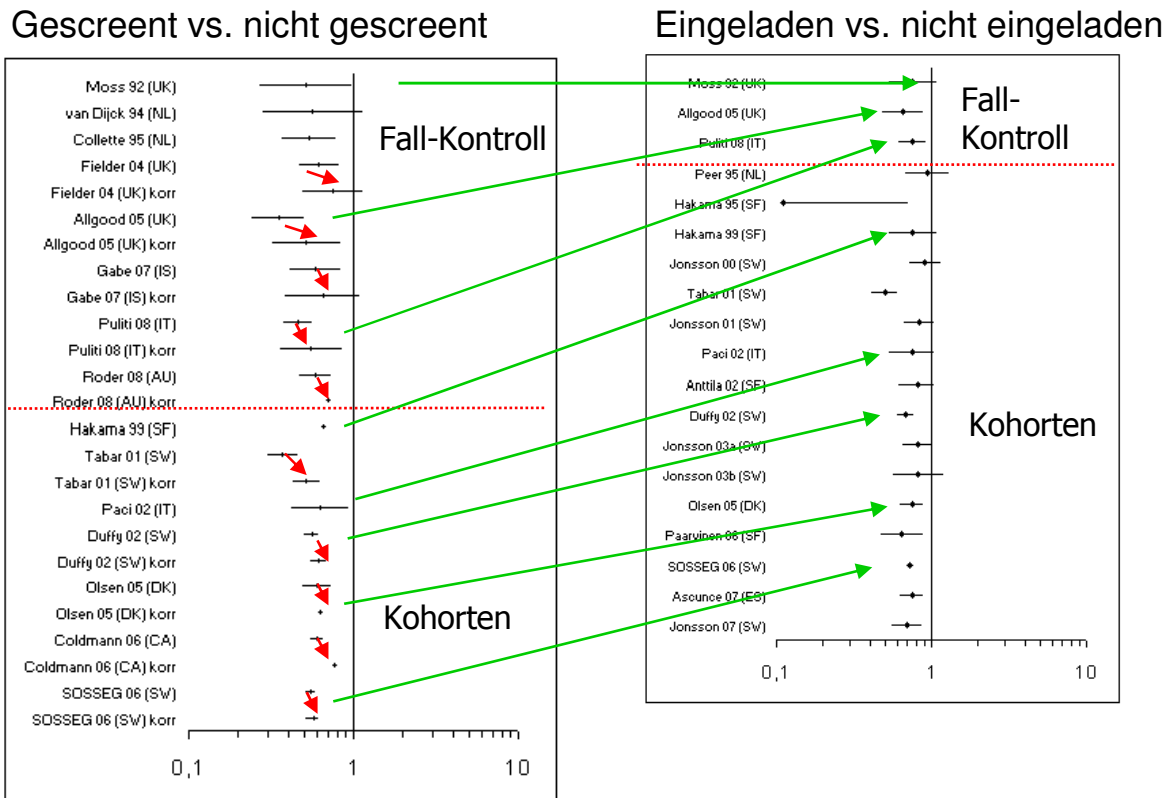


Abbildung 9. Veränderung der Ergebnisse je nach Definition der Exposition (grüne Pfeile) bzw. nach Korrektur für Self-Selection (rote Pfeile) (Quelle: eigene Darstellung mit Angaben der Studien)

4. Zusammenfassende Diskussion

In der vorliegenden Arbeit wurden insgesamt 40 Publikationen in Detail analysiert. Ziel der Arbeit war die Identifikation und Zusammenfassung von Strategien zur Umgang mit Verzerrungsquellen in der Evaluation von Screeningprogramme. Wie in jeder systematischen Übersicht, kann nicht ausgeschlossen werden, dass relevante Publikationen nicht identifiziert und nicht eingeschlossen wurden. Ein Vergleich mit einem aktuellen systematischen Review, der sich mit der Effektivität von Mammographie-Screening-Programmen befasste (1), zeigt, dass die uns vorliegenden Arbeiten die relevantesten Programme abdecken. Allerdings identifizierte der Review von Schopper und Wolf sechs zusätzliche Veröffentlichungen. Dabei handelt es sich zum Teil (n=3) um Berichte, die nicht als Zeitschriftenaufsatz veröffentlicht wurden und deshalb nicht durch die bibliographische Datenbankrecherche identifiziert werden konnten. Diese Berichte befassen sich mit den Programmen in Australien, England und den Niederlanden, die durch die uns vorliegende Literatur abgedeckt waren. Die restlichen 3 Veröffentlichungen sind Zeitschriftenaufsätze, die über Programme in Finnland und den Niederlande berichten, Länder die in unserem Studiensample auch abgedeckt waren. Es ist nicht auszuschließen, dass in diesen 6 Veröffentlichungen weitere Strategien zur Berücksichtigung von Bias berichtet werden. Eine schnelle Durchsicht⁸ dieser Veröffentlichungen lässt uns jedoch einschätzen, dass durch den Einschluss dieser Veröffentlichungen in unserer Arbeit keine weiteren Strategien identifiziert worden wären.

Ziel dieser Arbeit war die Identifizierung und Beschreibung verschiedene Strategien zur Umgang mit Bias bei der Evaluation von Mammographie-Screening-Programmen. Dies erfolgte anhand eines Studienpools, der überwiegend aus Kohorten- und Fall-Kontroll-Studien bestand. Insbesondere vier Biasarten werden in der Literatur im Kontext der Evaluation von Screening-Programme benannt: Lead-Time, Self-Selection, Over-Diagnosis und Length-Time (siehe Kapitel 1 Einführung). Die überwiegende Mehrheit der eingeschlossenen Studien berichtet explizit oder implizit über Strategien zur Berücksichtigung von Self-Selection- und Lead-Time-Bias. Im Gegensatz dazu werden die verwandten Over-Diagnosis- und Length-Time-Bias in weniger Studien adressiert. Darüber hinaus wurden in den vorliegenden Studien weitere potentielle Verzerrungsquellen beschrieben. Dabei wurden ähnliche Verzerrungsquellen von verschiedenen Autoren unterschiedlich bezeichnet, so dass die Terminologie nicht konsistent war.

In den hier vorliegenden Studien wurden verschiedene Methoden beschrieben um mit den multiplen Verzerrungsquellen umzugehen. Grundsätzlich lassen sich diese in zwei Gruppen unterteilen:

- 1) Strategien im Studiendesign und -durchführung, bei denen im Vorfeld versucht wird die potentiellen Verzerrungsquellen zu minimieren. Durch die Auswahl des Studiendesigns, der relevanten Endpunkte, durch die Definition von Exposition, Fälle, Kontrollen und Beobachtungszeiträume kann der potentiellen Einfluss der verschiedener Biasarten reduziert werden. Wählt man z.B. Brustkrebsmortalität als Outcomeparameter, ist die

⁸ Im gegebenen zeitlichen Rahmen des Projektes war es nicht möglich, zusätzliche Publikationen ausführlich zu bewerten. Eine Veröffentlichung aus den Niederlanden war nicht zugänglich.

Relevanz von Lead-Time geringer als wenn Überlebenszeit betrachtet wird (50). Wird die Exposition als erfolgte Mammographie definiert, erhöht sich das Risiko von Self-Selection-Bias.

- 2) Strategien zu Korrektur/ Adjustierung der Ergebnisse. Die Ergebnisse der Studien werden korrigiert, um Effekten potentieller Verzerrungsquellen Rechnung zu tragen – oder adjustiert, um den Effekt potentieller Störfaktoren zu kontrollieren.

Die Validität der verschiedenen Ansätze (sowohl im Studien-Design als auch zur Korrektur) hängt damit zusammen, in wie weit die getroffenen Annahmen selbst valide sind. Insgesamt beruhen alle hier identifizierte Ansätze zur Berücksichtigung von Self-Selection-Bias auf der plausiblen Annahme, dass dieser Bias eher zu einer Überschätzung der Effekte vom Screening führen kann (52). Neben dieser Grundannahme werden bei der Korrekturansätze für Self-Selection-Bias weitere Annahme zur Teilnehmerate und Mortalitätsrisiko bei Nicht-Teilnehmerinnen getroffen. Diese variieren zwischen den Studien. Die Annahmen zur Teilnehmerate lagen zwischen 65% und 85%. Die Annahmen zur Erhöhung des Mortalitätsrisiko bei Nicht-Teilnehmerinnen lagen in den Studien zwischen 11% und 36% (siehe Tabelle 8).

Bei der Analysen zur Überlebenszeit wurde die auch plausible Annahme getroffen, das dieser Bias zu einer Überschätzung der Effekte führen kann (50,53). Dabei variierte die angenommene Lead-Time zwischen 2,5 und 5 Jahren. Für Lead-Time wurde auch in manchen Studien mit dem Endpunkt Brustkrebsmortalität korrigiert, wobei je nach Studie gegenteilige Annahmen zum Effekt dieses Bias getroffen wurden. In zwei Studien wurde angenommen, dass Lead-Time-Bias zur Überschätzung der Effekte führen würde (27,35,36). Einige Studien (28,29,30,31,32) führten jedoch eine Korrektur der Ergebnisse zu Brustkrebsmortalität durch, die auf der Annahme beruhte, dass Lead-Time-Bias zu einer Unterschätzung der Effekte führen würde. Aus unserer Sicht wurden die jeweiligen Annahmen für die spezifische Studiensituation plausibel erklärt (siehe Kapitel 3.4.1).

Eine umfassendere systematische Bewertung der Validität und der Vor- und Nachteile der hier vorgestellten verschiedenen Strategien war in der vorliegenden Arbeit nicht möglich. Gerade beim Mammographie-Screening ist zu erwarten, dass die Validität, Angemessenheit oder Plausibilität der Annahmen verschiedener Aspekte des Studiendesigns kontrovers diskutiert werden. Im Zusammenhang mit der Veröffentlichung des Cochrane-Reviews der RCTs zu Mammographie-Screening entfachte sich z.B. eine kontroverse Diskussion über die Eignung von Brustkrebsmortalität als Endpunkt für die Bewertung des Nutzens dieser Intervention. Die vorliegenden Studien spiegeln jedoch diese methodologische Diskussion kaum wider. Zwar wird teilweise durch Verweise auf weitere Publikationen die Kontroverse um den Nutzenparameter erwähnt (z.B. in 33,34), eine ausführliche Darstellung der Argumente fehlt jedoch. Dieses Beispiel zeigt, dass für eine umfassende und systematische Bewertung der Validität der hier zusammengefassten Strategien bzw. Studien-Design-Merkmale die Einbeziehung weitere Publikationen erforderlich ist. Dies war aber unter den Rahmenbedingungen dieses Projektes jedoch nicht möglich. In der vorliegenden Arbeit wurden nur Publikationen eingeschlossen, die Evaluationen von Mammographie-Screening-Programmen berichteten. Theoretische bzw. methodologische Publikationen, die z.B. die Validität, Angemessenheit, oder die Vor- und Nachteile verschiedener Design-Strategien zur Berücksichtigung von Bias diskutiert hätten,






wurden nicht eingeschlossen. Aus diesem Grund war es auch in dieser Arbeit nicht möglich, die in den Studien berichteten Strategien mit Empfehlungen aus der methodologischen Literatur zu kontrastieren. Eine solche Gegenüberstellung von methodologischen Empfehlungen und Praxis wurde in einer Review aus dem Jahr 1998 für Fall-Kontroll-Studien zu Krebs-Screening unternommen (46). Dieser Review schloss 3 Fall-Kontroll-Studien ein, die in unserem Studiensample auch eingeschlossen waren. Die Autoren kamen zu der Schlussfolgerung, dass die Studien mit den methodologischen Empfehlungen konsistent bezüglich der Auswahl der Grundbevölkerung, der Auswahl der Fälle und der Kontrollen, und dem Umgang mit Self-Selection-Bias waren (46). Bei der Definition der Exposition fanden sie jedoch Diskrepanzen zwischen Theorie und Praxis (46). Aus unserer Sicht folgen die Auswahl der Grundbevölkerung, der Fälle und der Kontrolle in den restlichen Fall-Kontroll-Studien unseres Samples auch plausiblen Überlegungen. Bei der Definition der Exposition wurde überwiegend die Teilnahme am Screening verwendet. Die Definition der Exposition aufgrund des Einladungsstatus ist jedoch geeigneter, um Self-Selection-Bias zu berücksichtigen.

Die Evaluation von Mammographie-Screening-Programme erfordert die sorgfältige Berücksichtigung potentieller Verzerrungsquellen, unabhängig davon welches Studiendesign gewählt wird. In dieser Arbeit wurden verschiedene Strategien zur Umgang mit Bias, die beim Design einer Evaluation des Mammographie-Screenings außerhalb von RCTs verwendet werden können, diskutiert.

5. Extraktionstabellen

In diesem Kapitel befinden sich die Extraktionstabellen mit den relevanten Angaben aus der vom Auftraggeber übermittelten Publikationen. Die Tabellen fassen die wichtigsten Merkmale der Mammographie-Screening-Programme sowie die relevanten Studiendesign-Charakteristika zusammen.

Die Anordnung der Tabellen folgt in erster Linie dem Studiendesign und innerhalb des Studiendesigns den Ländern. Folgende Farben kennzeichnen die verschiedenen Studiendesigns:

-  Fall-Kontroll(-Studie)
-  Kohorten
-  Ökologisch
-  Quasi-Randomisiert
-  Fall-Kohorte

5.1 Charakteristika der Screening-Programme

Tabelle 10. Programm Charakteristika.

Studie	Programm	Land	Region (Zielbevölkerung*)	Programm Charakteristika					
				Beginn	Intervall	Alter	Teilnehmerinnen ^a	Aufnahme ^b	Befundung ^c
Roder 2008 ²	BreastScreen-SA	Australien	South Australia (k.A.)	1989 ^d / 1994 ^e	2 J	(40-49)-50-69-(>70) ^f	k.A.	2-Ebenen	2-fach
Allgood 2008 ⁴⁰	NHSBSP-EA	Großbritannien	East Anglia (k.A.)	1989	3 J	50-70	k.A.	k.A.	k.A.
Fielder 2004 ⁴¹	Breast Test Wales	Großbritannien	Wales (k. A.)	1989/1991 ^e	3 J	50-64	250000/round	k.A.	2-fach
Moss 1992 ⁴²	TEDBC	Großbritannien	Guildford/Stoke (70971)	1979 ^g	2 J	45-64	72%	k.A.	k.A.
Gabe 2007 ¹²	IPBMSP	Island	National (Island) (40000)	1987/1989 ^e	2 J	40-69	ca. 62%	2-Ebenen ^h	2-fach
Puliti 2008 ¹⁷	IMPACT	Italien	Regionen Toskana, Piedmont, Emilia-Romagna, Umbrien, Veneto (k.A.)	1990/1992/1995/1997/1999 ⁱ	2 J	50-69	k.A.	2-Ebenen	k.A.
Broeders 2002 ²⁰	Nijmegen	Niederlande	Nijmegen (30000)	1975/1997 ^j	2 J	35-74	k.A.	2-Ebenen/ 1-Ebene	1-fach
Collette 1984 ^{21,22,23,24}	DOM Projekt	Niederlande	Utrecht (20555)	1974	1-4 J	50-64	72%	2-Ebenen, 1-Ebene	k.A.
van Dijk 1995 ²⁶	Nijmegen	Niederlande	Nijmegen (10000)	1975/1977 ^j	2 J	35-69 (74 ^k) (>65 ^l)	k.A.	1-Ebene	k.A.
Spillane 2001 ³	BreastScreen-Australia	Australien	National (Daten nur aus einem Zentrum) (k.A.)	1991	2 J	(40-49)-50-69-(>70) ^f	40%-54%	2-Ebenen	k.A.
Olsen 2005 ^{5,6}	Kopenhagen	Denmark	Kopenhagen Gemeinde (40000)	1991	2 J	50-69 ^m	ca. 70%	2-Ebenen	2-fach
Anttila 2002 ⁷	Helsinki	Finnland	Stadt Helsinki (k.A.)	1986	2 J	50-59	144409 / 82%	2-Ebenen	2-fach
Parvinen 2006 ⁸	Cancer Society Finnland	Finnland	Helsinki, Turku Tampere (k.A.)	1987	2 J	50-59 (-69) ⁿ	k.A.	2-Ebenen	2-fach

Studie	Programm	Land	Region (Zielbevölkerung*)	Programm Charakteristika					
				Beginn	Intervall	Alter	Teilnehmerinnen ^a	Aufnahme ^b	Befundung ^c
Paci 2008 ¹⁵	Italien	Italien	Regionen Torino, Parma, Reggio-Emilia, Ferrara, Modena, Bologna, Romagna, Firenze, Perugia (k.A.)	k.A.	2 J	50-69	k.A.	2-Ebenen	k.A.
Paci 2002 ¹⁶	Florenz	Italien	Stadt Florenz (k.A.)	1990	2 J	50-69	ca. 71000 (% k.A.)	2-Ebenen	k.A.
Coldman 2006 ⁴	SMPBC	Kanada	British Columbia (k.A.)	1988	1 J (<50), 2 J	40-79	598690 / ca. 50%	k.A.	k.A.
Peer 1995 ²⁵	Nijmegen	Niederlande	Nijmegen (13500)	1975	2 J	35-65 (74 ^k) (35-49 ^l)	65% (87% ^o)	1-Ebene	k.A.
Duffy 2002 ²⁷	SW-7-Co	Schweden	7-Counties (k.A.)	1978-1994 ^p	1,5-2,75 J ^p	40-69/40-74/50-69 ^p	64%-92% ^p	k.A.	k.A.
Jonsson 2007 ²⁸	Nord-Schweden	Schweden	4 Counties (185000)	1989-1995 ^p	20-22 Monate	40-74	78%-88%	2-Ebenen	2-fach
Jonsson 2001 ²⁹	Schweden	Schweden	20 Counties (260594)	1986-87 ^p	23 Monate (Durchschnitt)	50-69 ^q	k.A.	k.A.	k.A.
Jonsson 2000 ³⁰	Schweden	Schweden	26 Counties (439431)	1986-1997 ^p	18-24 Monate (Durchschnitt 20 Monate)	40-49 ^q	k.A.	k.A.	k.A.
Jonsson 2003a ³¹	SW-Gav	Schweden	Gävlaborg County (43749)	1974/1979 ^e	2-2,5 J	40-74 (40-64 ^k)	84%	1-Ebene	k.A.
Jonsson 2003b ³²	Schweden	Schweden	23 Counties (125438)	1986-1990 ^p	23 Monate (Durchschnitt)	40-74 (70-74 ^k)	k.A.	k.A.	k.A.
SOSSEG 2006 ^{33,34}	Schweden	Schweden	13 Counties (1108610)	1980-1990 ^p	k.A.	40-69 / 50-69 ^p	70%-90% ^p	k.A.	k.A.
Tabar 2001 ^{35,36}	Two-County	Schweden	2 Counties (Dalarna/Östergötland) (504157)	1978	k.A.	40-69 / 40-74	85%	k.A.	k.A.
Ascunce 2007 ³⁸	Navarra	Spanien	Navarra (k.A.)	1990	2 J	45-65(69) ^f	k.A.	k.A.	k.A.
Quinn 1995 ⁴³	NHSBSP	Großbritannien	England und Wales (k.A.)	1988	2-3J ^s	50-64	k.A.	1-2 Ebenen ^s	1-2fach ^s
Barrichelli 2001 ¹³	Florenz	Italien	Region Florenz (118000)	1970-1995 ^t	2 J	50-69	ca. 60%	2-Ebenen	k.A.
Gorini	Florenz	Italien	Region Florenz (k.A.)	1970	2 J	40-69 / 50-69 ^u	ca. 28.000/J (% k.A.)	2-Ebenen	k.A.

Studie	Programm	Land	Region (Zielbevölkerung*)	Programm Charakteristika					
				Beginn	Intervall	Alter	Teilnehmerinnen ^a	Aufnahme ^b	Befundung ^c
2004 ¹⁴									
Törnberg 1994 ³⁷	Schweden	Schweden	26 Counties (k.A.)	1986	k.A.	50-74 ^q	k.A.	k.A.	k.A.
Hakama 1999 ^{9,10}	Cancer Society Finnland	Finnland	386 Kommunen (ca. 85% der finnischen Kommunen) (k.A.)	1987	2 J	50-59	76389 (85%)	2-Ebenen	2-fach
Hakama 1995 ¹¹	Kotka Pilot ^v	Finnland	Kotka City und Umgebung (k.A.)	1982	2 J	40-47	ca. 16000 (88%)	1-Ebene	1-(2-3)-fach
Thompson 1994 ⁴⁴	Puget-Sound Health Cooperative	USA	Western Washington State (94656)	1985	1-3 J	>=40	k.A.	k.A.	k.A.

ASA: Southern Australia; IPBMSP: Icelandic Population-Based Mammography Screening Program; J: Jahre; k.A.: keine Angaben; NHSBSP-EA: National Health Service Breast Screening Programme East Anglia; SMPBC: Screening Mammography Programme of British Columbia; SOSSEG: Swedish Organized Service Screening Evaluation Group. TEDBC: Trial of Early Detection of Breast Cancer.

*Zielbevölkerung: Anzahl der Frauen im Screeningalter im Studiengebiet.

^aAnzahl der Teilnehmerinnen seit Einführung/ Teilnehmerate der Zielbevölkerung; ^bRöntgen-Aufnahme; ^cAnzahl und Modus der Auswertung der Mammographie-Aufnahmen; ^dPilotprojekt; ^eZeitpunkt der flächendeckende Implementierung; ^fAktive Einladung von Frauen 50-69, andere Altersgruppen erlaubt. ^gAnfang des Trials. ^hSeit 1995 1-Ebene für sehr voluminösen Brüsten. ⁱje nach Region. ^jZeitpunkt der Erweiterung. ^kSpäter bis 74 erweitert. ^lIn der Studie analysierte Altersgruppe. ^mBeim 2. Screening wurde Frauen bis 71 eingeladen. ⁿWeitere Screening bis 69 J empfohlen. ^oIn der ersten Screeninground. ^pUnterschiedlich je nach County. ^qIn jeder County verschiedene Altersgruppen, Altersgruppe der Studie. ^rSeit 1998. ^sRegionale Unterschiede. ^tProgressive Implementierung des Programmes. ^uab 1990. ^vNeben Mammographie, das Screening-Program beinhaltete Anleitung zur Selbstersuchung und klinische Untersuchung durch Krankenschwester. ^wAnalyse der gängigen Praxis ohne organisiertes Screening-Programm.

5.2 Charakteristika der Studien

Tabelle 11. Studiendesign, Datenerhebung.

Studie	Programm	Studien-Design	Intervention Gruppe	Kontroll Gruppe	Endpunkte*	Datenquellen [Verlinkung]	Erhebungszeitraum (Follow-up)
Roder 2008 ²	BreastScreen-SA	Fall-Kontroll	Screening Teilnahme (nach Datenbank des Programmes)	Keine Teilnahme	BK-Mortalität	SA Cancer Registry Totenschein+Klinische Unterlagen [Geburtsdatum und PLZ]	1994-2005
Allgood 2008 ⁴⁰	NHSBSP-EA	Fall-Kontroll	Screening Teilnahme (mindestens einmal)	Keine Teilnahme	BK-Mortalität	EA Cancer Registry NHS-Database [nicht genau beschrieben]	1995-2004
Fielder 2004 ⁴¹	Breast Test Wales	Fall-Kontroll	Screening Teilnahme	Keine Teilnahme	BK-Mortalität	Breast Test Wales Datenbank Patient Episode Datenbank Krankenakten Krebsregister (Todesursache BK wenn im Totenschein als primäre oder sekundäre Ursache kodiert). [nicht genau beschrieben]	1998-2001
Moss 1992 ⁴²	TEDBC	Fall-Kontroll	Gescreent	Nicht-gescreent	BK-Mortalität	National Health Service Central Register	1979-1986
Puliti 2008 ¹⁷	IMPACT	Fall-Kontroll	Gescreent // Eingeladen	Nicht gescreent // nicht eingeladen	BK-Mortalität	Regionalen Mortalitätsregister Regionale Krebsregister Regionale Screening-Programme Datenbanken [nicht genau beschrieben]	1988-2002
Gabe 2007 ¹²	IPBMSP	Fall-Kontroll	Screening Teilnahme (mindestens einmal)	Keine Teilnahme	BK-Mortalität	National Register	1990-2002
Broeders 2002 ²⁰	Nijmegen	Fall-Kontroll	Gescreent [Teilnahme an Jüngste Einadung]	Nicht gescreent [keine Teilnahme bei jüngste Screeningeinladung]	BK-Mortalität	Nijmegen Screeningregister Nijmegen BK-Register Nijmegen Einwohnerregister Krankenakten [vermutlich Name]	1975-1997
Collette	DOM Projekt	Fall-Kontroll	Gescreent	Nicht gescreent	BK-Mortalität	ad hoc DOM-Krebsregister	1973-1981/1988/1992 ^b

Studie	Programm	Studien-Design	Intervention Gruppe	Kontroll Gruppe	Endpunkte*	Datenquellen [Verlinkung]	Erhebungszeitraum (Follow-up)
1984 ^{21,22,23,24}					Überleben ^a	Melderegister Utrecht Krankengeschichte/akten (aus GP, Krankenhaus) <i>[Name und Geburtsdatum]</i>	
van Dijk 1995 ²⁶	Nijmegen	Fall-Kontroll	Gescreent [Teilnahme in jüngste Screeningrunde + 1 negative Screening in vorletzter Runde]	Nicht gescreent [Kein Screening in die 5. Runde vor Indexrunde]	BK-Mortalität	Nijmegen Screeningregister Nijmegen BK-Register Nijmegen Einwohnerregister Krankenakten <i>[vermutlich Name]</i>	1977-1994
Spillane 2001 ³	BreastScreen Australia	Kohorten	BK entdeckt bei BreastScreen Australia	Alle andere BK	[Umfang der Operation Histologie Grading Tumogrösse Staging] Überleben	Klinische Register (Brustkrebs- Referenzzentrum)	1991-1997
Olsen 2005 ^{5,6}	Kopenhagen	Kohorten	Einladung zur Screeningteilnahme [Region Kopenhagen, Screeningperiode]	Keine Einladung zur Screening [Regionen ohne Screening in DK, Pre- Screeningperiode]	BK-Mortalität Gesamt mortalität	kommunales Mammographie-Screening- Register Dänisches zentrales Bevölkerungs- register Dänisches Krebsregister Register der dänischen Brustkrebs- kooperative <i>[Persönliche Identifikations-Nr.]</i>	1991-2001 1981-1991
Anttila 2002 ⁷	Helsinki	Kohorten	Eingeladen [Geburtskohorte 1935- 1939]	Vor Screeningeinführung (nicht eingeladen) [Geburtskohorte 1930- 1934]	BK-Mortalität (SMR)	Finnish Cancer Registry Mortalitätsregister (National Population Register) Helsinki Screening Register <i>[nicht genau beschrieben]</i>	1970-1998
Parvinen 2006 ⁸	Cancer Society Finnland	Kohorten	Eingeladen [Geburtskohor- ten 1918-1932, Frauen die in Screeningzeitraum (1987-1997) 55-69 J alt waren]	Vor Screeningeinführung (nicht eingeladen) [Geburtskohorte 1907- 1921, Frauen die in Pre- Screeningzeitraum (1976- 1986) 55-69 J alt waren]	BK-Mortalität	Finnish Cancer Registry Mortalitätsregister Screening Register <i>[nicht genau beschrieben]</i>	1987-2001 (Screen) 1976-1991 (nicht Screen)

Studie	Programm	Studien-Design	Intervention Gruppe	Kontroll Gruppe	Endpunkte*	Datenquellen [Verlinkung]	Erhebungszeitraum (Follow-up)
Paci 2008 ¹⁵	Italien	Kohorten	Eingeladen [bzw. Nach Einführung des Programms]	Nicht eingeladen [vor Einführung des Programms]	Überleben	Regionale Mortalitätsregistern Census Regionale Krebsregister <i>[nicht genau beschrieben]</i>	1 Jahr vor Initiierung des Programmes bis 2005 (10 J)
Paci 2002 ¹⁶	Florenz	Kohorten	Gescreent // Eingeladen	nicht gescreent / nicht eingeladen [vor Einführung des Programms]	BK-Mortalität	Tuscany Cancer Registry	1990-1999
Coldman 2006 ⁴	SMPBC	Kohorten	Screening Teilnahme (Selbst-Referral)	Keine Teilnahme	BK-Mortalität	BC Cancer Registry Vital Statistics (Totenschein) <i>[Persönliche Identifikations-Nr, Adresse, Name, Geburtsdatum mit Automatch-Software]</i>	1988-2003
Peer 1995 ²⁵	Nijmegen	Kohorten	Eingeladen [Nijmegen Program]	Nicht eingeladen [Arnhem, kein Screeningprogramm]	BK-Mortalität	Nijmegen Screeningregister Nijmegen BK-Register Nijmegen Einwohnerregister Central Dutch Bureau of Statistics Krankenakten <i>[vermutlich Name]</i>	1975-1990
Duffy 2002 ²⁷	SW-7-Co	Kohorten	Eingeladen [Zeitraum nach Einführung des Screenings] Tatsächlich gescreent [Zeitraum nach Einführung des Screenings]	Nicht Eingeladen [Zeitraum vor Screening-Einführung] Eingeladen, aber nicht gescreent [Zeitraum während Screening]	BK-Mortalität	Regional Oncologic Centers Statistics Sweden (Mortalitätsregister) Screening-Zentren <i>[nicht genau beschrieben]</i>	Unterschiedlich je nach County
Jonsson 2007 ²⁸	Nord-Schweden	Kohorten	Eingeladen [Counties mit Screeningprogramm]	Nicht Eingeladen [Counties ohne Screeningprogramm]	Excess BK-Mortalität (siehe Seite 18) BK-als Todesursache	Regional Cancer Register Statistics Sweden (Mortalitätsregister) Screening Zentren <i>[nicht genau beschrieben]</i>	1989-2001 (11 Jahre)
Jonsson 2001 ²⁹	Schweden	Kohorten	Eingeladen [Frauen in Counties mit Screening Start bei 1986-87]	Nicht Eingeladen [Frauen in Counties die erst ab 1993-97 Screening hatten]	Excess BK-Mortalität (siehe Seite 18) BK-als Todesursache	Swedish Cancer Registry Statistics Sweden (Mortalitätsregister) Screening Zentren <i>[nicht genau beschrieben, Screeninggeschichte nicht auf individuelle Ebene verlinkt]</i>	1986-1997 (11 Jahre)

Studie	Programm	Studien-Design	Intervention Gruppe	Kontroll Gruppe	Endpunkte*	Datenquellen [Verlinkung]	Erhebungszeitraum (Follow-up)
Jonsson 2000 ³⁰	Schweden	Kohorten	Eingeladen [Frauen in Counties die 40-49 J alt Frauen einladen]	Nicht Eingeladen [Frauen in Counties die nur ab 50 J einladen]	BK-Mortalität	Swedish Cancer Registry Statistics Sweden (Mortalitätsregister) Screening Zentren <i>[nicht genau beschrieben, Screeninggeschichte nicht auf individuelle Ebene verlinkt]</i>	1986-1996 (8 Jahre)
Jonsson 2003a ³¹	SW-Gav	Kohorten	Eingeladen [Gävlaborg County]	Nicht eingeladen [Nachbar Counties / Rest Schweden (mit Ausnahme der Counties die an RCT teilnahmen)]	Excess BK-Mortalität (siehe Seite 18) BK-als Todesursache	Swedish Cancer Registry Statistics Sweden (Mortalitätsregister) Screening Zentren <i>[nicht genau beschrieben, Screeninggeschichte nicht auf individuelle Ebene verlinkt]</i>	1976-1998 (22 Jahre)
Jonsson 2003b ³²	Schweden	Kohorten	Eingeladen [Frauen in Counties, die bis 74 J einladen]	Nicht Eingeladen [Frauen in Counties, die nur bis 69 J einladen]	Excess BK-Mortalität (siehe Seite 18) BK-als Todesursache	Swedish Cancer Registry Statistics Sweden (Mortalitätsregister) Screening Zentren <i>[nicht genau beschrieben, Screeninggeschichte nicht auf individuelle Ebene verlinkt]</i>	1986-1998 (10 Jahre)
SOSSEG 2006 ^{33,34}	Schweden	Kohorten	Eingeladen [Zeitraum nach Einführung des Screenings] Tatsächlich gescreent [Zeitraum nach Einführung des Screenings]	Nicht Eingeladen [Zeitraum vor Screening-einführung] Eingeladen, aber nicht gescreent [Zeitraum während Screening]	BK-Mortalität	Regional Oncologic Centers Statistics Sweden (Mortalitätsregister) Screening Zentren <i>[nicht genau beschrieben]</i>	Unterschiedlich je nach County (13 Jahre)
Tabar 2001 ^{35,36}	Two-County	Kohorten	Eingeladen [Zeitraum nach Einführung des Screenings] Tatsächlich gescreent [Zeitraum nach Einführung des Screenings]	Nicht eingeladen [Zeitraum vor Screening-einführung] Eingeladen, aber nicht gescreent [Zeitraum während Screening]	BK-Mortalität, Überleben	Regional Oncologic Centers Swedish Cancer Registry Statistics Sweden (Mortalitätsregister) <i>[nicht genau beschrieben]</i>	1978-1996 (Screen) 1968-1977 (nicht Screen)
Ascunce 2007 ³⁸	Navarra	Kohorten	Nach der Einführung von Screening (1997-2001)	Vor der Einführung von Screening (1986-1990)	BK-Mortalität	Navarre Todesursachen-Register (Register der Todesfälle in der Region und Einwohner die woanders gestorben sind). Mortalität in Spanien Navarre Krebsregister	1975-2004

Studie	Programm	Studien-Design	Intervention Gruppe	Kontroll Gruppe	Endpunkte*	Datenquellen [Verlinkung]	Erhebungszeitraum (Follow-up)
						[nicht genau beschrieben]	
Quinn 1995 ⁴³	NHSBSP	Ökologische Studie	Nach der Einführung des Programms (1988-1994)	Vor der Einführung des Programms (1950-1987)	BK-Mortalität	Regionale Krebsregister Census-Register	1950-1994
Barrichelli ¹³	Florenz	Ökologische Studie	Exponiert zu Screening-Programm [Gebiet mit Screening]	Nicht exponiert zu Screening-Programm [Gebiete ohne Screening]	BK-Mortalität BK-Inzidenz 5J-Überleben	National Statistics Bureau Tuscany Regional Mortality Registry, Tuscany Cancer Registry	1970-1997 / 1985-86 vs. 1991-92 ^d
Gorini 2004 ¹⁴	Florenz	Ökologische Studie	Gebiete mit Screening seit 1970	Gebiete mit Screening seit 1990	BK-Mortalität	Tuscany Mortality Register Tuscany Cancer Register	1985-2000
Törnberg 1994 ³⁷	Schweden	Ökologische Studie	Counties mit Screening	Counties ohne Screening	BK-Mortalität	Swedish Cancer Registry Statistics Sweden (Mortalitätsregister)	1971-1990
Hakama 1999 ^{9,10}	Cancer Society Finnland	quasi-randomisiert	Eingeladen (Gescreent) [gerade Jahre Geburtskohorte]	Nicht eingeladen [ungerade Jahre Geburtskohorte]	BK-Mortalität (SMR)	Finnish Cancer Registry Mortalitätsregister (National Population Register Centralised Mass Screening Register [nicht genau beschrieben]	1987-1992
Hakama 1995 ¹¹	Kotka Pilot	quasi-randomisiert	Eingeladen (Gescreent) [gerade Jahre Geburtskohorte]	Nicht eingeladen [ungerade Jahre Geburtskohorte]	BK-Mortalität (SMR)	Finnish Cancer Registry Mortalitätsregister Screening Register [nicht genau beschrieben]	1982-1991
Thompson 1994 ⁴⁴	Puget-Sound Health Cooperative	Fall-Kohorte	Screening Teilnahme	Keine Teilnahme	BK-Mortalität	Krebsregister (Ca. Surveillance System) Totenschein Klinische Akten [nicht genau beschrieben]	1982-1988

* Nur für die Evaluation der Effektivität des Programms relevante Endpunkte (z.B. BK-Mortalität). Viele Studien berichten auch andere Parameter (BK-Inzidenz, Detektionsrate, etc.), bei Studien die nur Surrogatendpunkte berichten, sind diese in [Klammern] gelistet worden.

BC: British Columbia; BK: Brustkrebs; GP: Hausarzt; NHSBSP-EA: National Health Service Breast Screening Programme East Anglia; SA: Southern Australia; SMR: Standardisierte Mortalitätsrate; TEDBC: Trial of Early Detection of Breast Cancer;

^aÜberleben würde nicht im Fall-Kontrolle Studiendesign sondern im Kohortendesign untersucht (Publikation 21) ^bJe nach Publikation. ^cIn jeder Region unterschiedliche Periode. ^dMortalitäts- und Inzidenzanalyse / Überlebensanalyse.

Tabelle 12. Umgang mit Bias.

Studie	Porgramm	Studien-Design	Strategien zur Umgang mit potentiellen Bias				
			<i>Lead-Time</i>	<i>Over-Diagnosis</i>	<i>Length-Time</i>	<i>Self-Selection</i>	<i>Andere</i>
Roder 2008 ²	BreastScreen-SA	Fall-Kontroll	Erwähnt das gesamte Studiendesign das Risiko von Lead-Time-Bias reduziert.	k.A.	k.A.	Korrektur der Ergebnisse bei einem Faktor von 1,19 (Duffy Methode, siehe Seite 26). Begleitende Survey über Faktoren die mit der Teilnahme an Screening zusammenhängen, als Interpretationshilfe: Ergebnis war das Teilnehmerinnen erhöhten Risiko hatte	<i>Time of Diagnosis</i> Ausschluss von Fällen, deren Diagnose vor der Implementierung des Programms lag (vor 1994, Zeitpunkt der vollen Implementierung). <i>Setup</i> Ausschluss von Fällen in der Anlaufphase (Analyse beginnt erst nach vollem Implementierung des Programs) <i>Screen Detected Cases</i> Ausschluss von BK-fällen die in den ersten drei Jahren diagnostiziert wurde <i>Exposure Opportunity</i> Ausschluss von Kontrollen mit BK-Diagnose, die vor respektiven Fall-BK-Diagnose liegt. <i>Missclassification Todesursache</i> Ermittlung der Todesursache verblindet für Screening-Status
Allgood 2005 ⁴⁰	NHSBSP-EA	Fall-Kontroll	k.A.	Einschluss von Diagnosen nach mind. 6 J nach Implementierung des Programmes (Ausschluss von prävalenten Erkrankungen)	k.A.	Korrektur der Ergebnisse mit einem Faktor von 1,86 (Duffy Methode, siehe Seite 26).	<i>Setup</i> Einschluss von Fällen deren Diagnose mindestens 6 J nach Implementierung des Programms
Fielder 2004 ⁴¹	Breast Test Wales	Fall-Kontroll	k.A.	k.A.	k.A.	Korrektur der Ergebnisse bei einem Faktor von 1,53 (Duffy Methode, siehe Seite 26).	<i>Time of Diagnosis</i> Ausschluss von Fällen, deren Diagnose vor der Implementierung des Programms lag. <i>Setup/Screen-Detected-Cases</i> Ausschluss aller BK-Diagnosen aus der ersten drei Jahre nach der Implementierung des Screenings
Moss 1992 ⁴²	TEDBC	Fall-Kontroll	k.A.	k.A.	k.A.	Erörterung des Verzerrungspotentials durch	-

Studie	Porgramm	Studien-Design	Strategien zur Umgang mit potentiellen Bias				
			Lead-Time	Over-Diagnosis	Length-Time	Self-Selection	Andere
						den Vergleich der BK-Mortalitätsrisikoreduktion wenn nur Screeningregion berücksichtigt wurde und wenn eine Region mit und eine ohne Screening berücksichtigt wurden (siehe Seite 27)	
Gabe 2007 ¹²	IPBMSP	Fall-Kontroll	k.A.	k.A.	k.A.	Korrektur der Ergebnisse bei einem Faktor von 1,27 (Duffy Methode, siehe Seite 26).	<i>Time of Diagnosis</i> Ausschluss von Fällen, deren Diagnose vor 1987/1990 lag oder <40J oder Geburtsdatum >=1919 oder starben vor 1990. <i>Exposure Opportunity</i> Sensitivitäts-Analyse mit Ein-/Ausschluss von Detektions-Screening und Korrektur der Ergebnisse nach der Methode von Duffy (siehe Seite 30).
Puliti 2008 ¹⁷	IMPACT	Fall-Kontroll	Nachlagerung der Pseudodiagnosidatum der Kontrolle (siehe Seite 20)	k.A.	k.A.	Korrektur der Ergebnisse mit einem Faktor von 1,19 (Duffy Methode, siehe Seite 26).	<i>Misclassification der Fälle</i> Ausmass abgeschätzt durch Vergleich der Ergebnisse unter Einschluss aller Todesursachen
Broeders 2002 ²⁰	Nijmegen	Fall-Kontroll	k.A.	k.A.	k.A.	Erwähnt Potential für Self-Selection-Bias, keine Korrektur.	<i>Setup</i> Betrachtung der 10 jüngsten Jahren des Programs <i>Misclassification Todesursache</i> Assessment der Todesursache verblindet für Screeningstatus.
Collette 1984 ^{21,22,23,24}	DOM Projekt	Fall-Kontroll	<i>bei Mortalität</i> Analyse unter Ausschluss der Fällen die kurz nach Diagnose starben (z.B. Ausschluss von Fällen die innerhalb des 1. J nach Diagnose starben) <i>bei Überleben</i> Überlebenszeit wurde bei Detektion-Screening um	k.A.	Erwähnt potential für Length-Time Bias in der Analyse des Überlebens ²¹	Erwähnt Potential für Self-Selection-Bias, keine Korrektur vorgenommen.	<i>Misclassification Todesursache</i> Überprüfung von Krankenakten und Ausschluss von Todesfälle mit Diskrepanzen zwischen Krankengeschichte und Totenschein. <i>Misclassification of Exposure</i> Sensitivitäts-Analyse mit Ein-/Ausschluss von Detektions-Screening

Studie	Porgramm	Studien-Design	Strategien zur Umgang mit potentiellen Bias					
			Lead-Time	Over-Diagnosis	Length-Time	Self-Selection	Andere	
			2,5 J (Lead-Time-Annahme) korrigiert und bei 0,5x Screeningintervall für die Detektion bei darauffolgenden Screening					
van Dijck 1995 ²⁶	Nijmegen	Fall-Kontroll	k.A.	k.A.	k.A.	Erwähnt Potential für Self-Selection-Bias, keine Korrektur. Abschätzung durch Vergleich der Überleben bei der nicht gescreenten mit der Überleben einer Nachbarstadt ohne Screening	<i>Misclassification Todesursache</i> Assessment der Todesursache verblindet für Screeningstatus.	
Spillane 2001 ³	Breast Screen Australia	Kohorten	Annahme eines Lead-Times von 3 J und entsprechend Vergleich der 5J-Überleben bei nicht gescreenten mit der 8J-Überleben bei gescreenten	k.A.	Höheren Anteil von Grad 1 Tumoren in der Screeninggruppe weist auf Length-Time-Bias hin	k.A.	k.A.	
Olsen 2005 ^{5,6}	Kopenhagen	Kohorten	k.A.	diskutiert Verzerrungspotential durch Overdetection-Bias, als nicht relevant eingeschätzt, da BK-Inzidenz unter Screening nicht anstieg	k.A.	diskutiert Verzerrungspotential durch sozioökonom. Faktoren, Alter, Region, Zeit als mgl. confounder berücksichtigt Subgruppenanalyse von am Screening teilnehmenden Frauen: Korrektur des RR mit einem Faktor für nichtteilnehmend	<i>Time of Diagnosis</i> Ausschluss von Frauen mit BK-Diagnose vor Screening <i>Misclassification</i> erwähnt aber als nicht relevant eingeschätzt, da RR~1 für andere Krebsarten. <i>Recall</i> erwähnt aber als nicht relevant eingeschätzt, da umfassende administrative Register Datengrundlage bilden. <i>Different follow up periods</i> gleiche follow up-Dauer für Studien- und Kontrollgruppen, Rekonstruktion eines Einladungssystems für Kontrollgruppen	

Studie	Porgramm	Studien-Design	Strategien zur Umgang mit potentiellen Bias				
			Lead-Time	Over-Diagnosis	Length-Time	Self-Selection	Andere
Anttila 2002 ⁷	Helsinki	Kohorten	k.A.	Observed/Expected BK Inzidenz bei Screening, schätzt Over-Diagnosis bei 18% (RR 1,18)	k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen.	<i>Time of Diagnosis</i> Ausschluss Todesfälle deren BK-Diagnose vor dem 50. Lebensjahr gemacht wurde.
Parvinen 2006 ⁸	Cancer Society Finnland	Kohorten	Erwähnt Lead-Time-Bias als relevant für Inzidenz in der Diskussion, aber keine Einschätzung	Erwähnt Over-Diagnosis, schätzt das es gering ist, da keine erhöhte Proportion von nicht invasiven Ca. <i>In Situ</i> Ca. wurden ausgeschlossen	k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen.	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden.
Paci 2008 ¹⁵	Italien	Kohorten	Annahme eines Lead-Times von 5 J und Analyse der Hazard Ratio in der Zeit danach (5. bis 10. J follow-up).	siehe Length-Time (Autoren geben an dass es sich hierbei um der extremen Form von Length-Time-Biashandelt).	Adjustierung für Tumor-Charakterisitka (Ausdehnung und Grading) bei Diagnose	k.A.	-
Paci 2002 ¹⁶	Florenz	Kohorten	Vergleich zwischen Pre-Screening- und Screeningzeitraum	Inzidenzbasierte Mortalitätsrate (siehe Seite 18)	Inzidenzbasierte Mortalitätsrate (siehe Seite 18)	k.A.	-
Coldman 2006 ⁴	SMPBC	Kohorten	k.A.	k.A.	k.A.	Abschätzung beim Vergleich der BK-Inzidenz in der Nicht-Screeninggruppe mit der Inzidenz in der Gesamtbevölkerung in der Zeit vor Screening-Einführung. Korrektur der Ergebnisse bei einem Faktor von 1,26 (Duffy Methode, siehe Seite 26)	<i>Setup</i> Ausschluss von BK-Todesfällen, deren Diagnose innerhalb der ersten 6 Monate nach Implementierung des Screenings gemacht wurde.
Peer 1995 ²⁵	Nijmegen	Kohorten		Erwähnt Potential für Overdetection. Berichtet höhere BK-Inzidenz in Interventionsgruppe und höhere Anteil von <i>in Situ</i> duktales Ca. (DCIS) in Interventionsgruppe			<i>Missklassifikation Todesursache</i> Gleiche Methode zur Assessment in Interventions und Kontrolregion

Studie	Porgramm	Studien-Design	Strategien zur Umgang mit potentiellen Bias				
			Lead-Time	Over-Diagnosis	Length-Time	Self-Selection	Andere
Duffy 2002 ²⁷	SW-7-Co	Kohorten	Korrektur der Personen Jahre in Screeninggruppe (Reduktion um die Anzahl von PJ, die aufgrund von Lead-Time nicht mehr im Interventionszeitraum zur Verfügung stehen; siehe Seite 21).	k.A.	k.A.	Korrektur der Ergebnisse nach der Duffy Methode (siehe Seite 26).	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden
Jonsson 2007 ²⁸	Nord-Schweden	Kohorten	Berechnet Überlebenunterschied (0,29 J länger beim Screeninggruppe) und ermittelt Todesfälle in diesem Zeitraum (2% aller BK-Todesfälle in Screeninggruppe) RR wird danach korrigiert. (siehe Seite 21).	Excess Mortalität	k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden.
Jonsson 2001 ²⁹	Schweden	Kohorten	Berechnet Überlebenunterschied (0,16 J länger beim Screeninggruppe) und ermittelt Todesfälle in diesem Zeitraum (0,4% aller BK-Todesfälle in Screeninggruppe) RR wird danach korrigiert. (siehe Seite 21).	Exzess Mortalität	k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen.	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden. <i>Setup</i> Korrektur der Ergebnisse (siehe Seite 30)
Jonsson 2000 ³⁰	Schweden	Kohorten	Erwähnt potential für Lead-Time-Bias (Überleben in gescreente war 0,11 Jahre länger). Würde zu geringfügige Unterschätzung (0,4%) von Screening führen (siehe Seite 21).	k.A.	k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen.	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden. <i>Setup</i> Korrektur der Ergebnisse (siehe Seite 30)
Jonsson	SW-Gav	Kohorten	Berechnet	Exzess Mortalität	k.A.	Vergleich zwischen eingeladen	<i>Time of Diagnosis</i>

Studie	Porgramm	Studien-Design	Strategien zur Umgang mit potentiellen Bias				
			Lead-Time	Over-Diagnosis	Length-Time	Self-Selection	Andere
2003a ³¹			Überlebenunterschied (1,07 J länger beim Screeninggruppe) und ermittelt Todesfälle in diesem Zeitraum (4,2% aller BK-Todesfälle in Screeninggruppe) RR wird danach korrigiert. (siehe Seite 21).			(unabhängig von Teilnahme) vs. nicht eingeladen.	Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden. <i>Setup</i> Korrektur der Ergebnisse (siehe Seite 30)
Jonsson 2003b ³²	Schweden	Kohorten	Berechnet Überlebenunterschied (0,1 J länger beim Screeninggruppe) und ermittelt Todesfälle in diesem Zeitraum (1,7% aller BK-Todesfälle in Screeninggruppe) RR wird danach korrigiert. (siehe Seite 21).	Exzess Mortalität	k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen.	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden. <i>Setup</i> Korrektur der Ergebnisse (siehe Seite 30)
SOSSEG 2006 ^{33,34}	Schweden	Kohorten	Korrektur Faktor in Poisson Regression Model. Follow-up Time korrigiert für Lead-Time	k.A.	k.A.	Korrektur der Ergebnisse nach Variation der Duffy Methode (siehe Seite 26).	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden.
Tabar 2001 ^{35,36}	Two-County	Kohorten	Annahme von 3-4 Jahre Lead Time, Einschluss von Todesfälle von Frauen deren BK 3-4 J vor Studienzeitraum diagnostiziert wurde in Interventionsgruppe (siehe Seite 22).	k.A.	k.A.	Korrektur der Ergebnisse (siehe Seite 26).	<i>Time of Diagnosis</i> Berücksichtigt nur BK-Fälle die in der „Screening Inzidenzzeitraum“ diagnostiziert wurden <i>Outcome</i> Vergleich der Ergebnisse zu BK-Mortalität mit den Ergebnisse von Gesamt mortalität bei Frauen mit BK und ohne BK
Ascunce 2007 ³⁸	Navarra	Kohorten	k.A.	k.A.	k.A.	k.A.	<i>Setup/Time of Diagnosis</i> Ausschluss der BK-Diagnosen die vor 1991 (voll-Implementierung) stattfanden. <i>Zeit-Trend</i>

Studie	Porgramm	Studien-Design	Strategien zur Umgang mit potentiellen Bias				
			<i>Lead-Time</i>	<i>Over-Diagnosis</i>	<i>Length-Time</i>	<i>Self-Selection</i>	<i>Andere</i>
							Vergleich zwischen Mortalitätsreduktion bei Gruppen die nicht an Screening teilnahmen (<49, >70) und Mortalitätsreduktion bei gescreente.
Quinn 1995 ⁴³	NHSBSP	Ökologische Studie	Erwähnt Möglichleit von Lead-Time-Bias	Erwähnt Erhöhung von Diagnosen in der ersten Jahren des Screenings	k.A.	k.A.	-
Barrichelli ¹³	Florenz	Ökologische Studie	k.A.	k.A.	k.A.	k.A.	-
Gorini 2004 ¹⁴	Florenz	Ökologische Studie	k.A.	k.A.	k.A.	k.A.	-
Törnberg 1994 ³⁷	Schweden	Ökologische Studie	k.A.	k.A.	k.A.	In den Mammographie-Programm-Extension-Score gingen die eingeladenen Frauen ein, unabhängig von Teilnahme (siehe Seite 30)	<i>Setup</i> Annahmen: Programm erreicht 50% des potentiellen Effekts zwischen 5-10 J nach Implementierung und 100% des pot. Effekts 10 J nach Implementierung. Dies wurde in der Mammographie-Programm-Extension-Score berücksichtigt (siehe Seite 30)
Hakama 1999 ^{9,10}	Cancer Society Finnland	quasi-randomisiert	k.A.		k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen.	<i>Time of Diagnosis</i> siehe SensitivitätsAnalyse
Hakama 1995 ¹¹	Kotka Pilot	quasi-randomisiert	k.A.	k.A.	k.A.	Vergleich zwischen eingeladen (unabhängig von Teilnahme) vs. nicht eingeladen.	<i>Time of Diagnosis</i> siehe SensitivitätsAnalyse
Thompson 1994 ⁴⁴	Puget-Sound Health Co-operative	Fall-Kohorte	Erwähnt das durch Fall-Kohortendesign und Auswahl von Mortalität nicht auftreten kann	k.A.	Erwähnt das durch Fall-Kohortendesign und Auswahl von Mortalität nicht auftreten kann	Multivariate Analyse mit Risikofaktoren	<i>Exposure Ascertainment</i> Analyse von Krankenakten

k.A: keine Angaben bedeutet NICHT EXPLIZIT ERWÄHNT; SA: Southern Australia

Tabelle 13. Analyse.

Studie	Programm	Studien-Design	Statistische Verfahren	Sensitivitätsanalyse ^a	Confounder / Effekt Modifier	
					In der Analyse	In der Diskussion
Roder 2008 ²	BreastScreen-SA	Fall-Kontroll	Logistische Regression		Alter (Matching) SES, Entfernung/Zugang zu Gesundheitsversorgung	Vorgeschichte (Brust-Operationen, Brustkrebs in der Familie, HRT)
Allgood 2008 ⁴⁰	NHSBSP-EA	Fall-Kontroll	Logistische Regression	Mit Korrektur nach Duffy oder mit Adjustierung für SES Diagnosen nach 1999	Alter (Matching) SES	k.A.
Fielder 2004 ⁴¹	Breast Test Wales	Fall-Kontroll	Logistische Regression für Matched Studien	Kontrolle aus der selben GP wie Fälle vs. aus anderen GP	Alter (Matching)	
Moss 1992 ⁴²	TEDBC	Fall-Kontroll	Mantel-Haenszel für matched Fall-Kontrolle Logistische Regression	Fälle und Kontrolle nur aus Screeningregion (Guildford) vs. Fälle und Kontrolle aus Screening und nicht-Screeningregion (Guildford bzw. Stoke)	Alter (Matching)	„Risiko Faktoren“, nicht näher spezifiziert, bis auf berufliche Status.
Gabe 2007 ¹²	IPBMSP	Fall-Kontroll	Logistische Regression	Ein-/Ausschluss von Detektion-Screening	Alter, Region (Matching)	k.A.
Puliti 2008 ¹⁷	IMPACT	Fall-Kontroll	Logistische Regression	Einschluss alle Todesursachen bei Brustkrebsfällen Verschiedene Expositionzeiten für Kontrollen (6 Monate, 1 J, 1,5 J)	Alter, Wohnort (Matching)	Verbesserte Therapie
Broeders 2002 ²⁰	Nijmegen	Fall-Kontroll		k.A.	k.A.	
Collette 1984 ^{21,22,23,24}	DOM Projekt	Fall-Kontroll	Konditional Logistische Regression für matched Studien. Überlebensanalyse mittels actuarial method	k.A.	Alter (Matching)	Verbesserte Therapie, Veränderte Aggressivität der Tumoren (Veränderte Prognose).
van Dijck 1995 ²⁶	Nijmegen	Fall-Kontroll	Konditional Logistische Regression für matched Studien	k.A.	Alter (Matching)	
Spillane 2001 ³	BreastScreen Australien	Kohorten	t-Test, Mann-Withney U, Chi-Quadrat,	k.A.	k.A.	<i>Dilution/Kontamination</i> De-facto Screening ausserhalb des Programs wird als nicht Screening klassifiziert und dadurch kann die

Studie	Programm	Studien-Design	Statistische Verfahren	Sensitivitätsanalyse ^a	Confounder / Effekt Modifier	
					In der Analyse	In der Diskussion
						Effektivität von Screening unterschätzt werden. Alter,
Olsen 2005 ^{5,6}	Kopenhagen	Kohorten	Poisson Regression	k.A.	Alter, Zeitperiode, Region	Regionale Unterschiede in Therapie, in Mortalitätsrisiko als nicht relevant eingeschätzt, da Diagnose und Therapie von Brustkrebs nach Guidelines in ganz Dänemark einheitlich erfolge
Anttila 2002 ⁷	Helsinki	Kohorten	Vergleich von BK-Mortalitätsrate	k.A.	Behandlungseffekte / Andere Hintergrundphänomene (Adjustierung für Sterblichkeit in der Pre-Screening Alter 40-49 J).	
Parvinen 2006 ⁸	Cancer Society Finnland	Kohorten	Vergleich von BK-Mortalitätsrate (95%-CI mittels Poisson)	k.A.	Behandlungseffekte / andre Hintergrundphänomene (Adjustierung für zeitliche Trends und Altersgruppen)	
Paci 2008 ¹⁵	Italien	Kohorten	Kaplan-Meier Survival Analyse mit Cox Model	k.A.	Alter, Tumor Stage (TN), Grading, Diagnose Jahr und Register	k.A.
Paci 2002 ¹⁶	Florenz	Kohorten	Vergleich der inzidenz-basierten Mortalitätsrate (siehe Seite 18)	k.A.	k.A.	Verbesserte Therapie
Coldman 2006 ⁴	SMPBC	Kohorten	Cox-Regression Observed/Expected BK-Todesraten (Poisson)	k.A.	Alter, Zeitpunkt der Diagnose, SES (Durchschnitt Einkommen im Wohnort)	Selbst-Einschätzung/Wahrnehmung der eigener Gesundheit
Peer 1995 ²⁵	Nijmegen	Kohorten	Vergleich cumulative Mortalitätsrate	k.A.	k.A.	Unterschiede zwischen beiden Regionen (Abschätzung der potentiellen Verzerrung durch Vergleich der BK-Mortalität in Periode von Screening).
Duffy 2002 ²⁷	SW-7-Co.	Kohorten	Log-lineare Regression	Eingeladen/nicht eingeladen vs. gescreent/nicht gescreent.	County	Verbesserung der Therapie, Aufmerksamkeit für BK Vergleiche zwischen tatsächlich gescreent und nicht gescreent, in der Screeningphase um sekuläre Trends auszuschließen. Abschätzung der

Studie	Programm	Studien-Design	Statistische Verfahren	Sensitivitätsanalyse ^a	Confounder / Effekt Modifier	
					In der Analyse	In der Diskussion
						zeitlichen Trends vor Screening.
Jonsson 2007 ²⁸	Nord-Schweden	Kohorten	Poisson Regression		Unterschiede in Hintergrund BK-Mortalität (Adjustierung mittels Pre-Screening Referenzgruppe in Screening und Kontrolregionen) Alter	
Jonsson 2001 ²⁹	Schweden	Kohorten	Multiplicative Poisson Regression	Exzess Mortalität vs. Mortalität von BK als Todesursache.	Unterschiede in Hintergrund BK-Mortalität (Adjustierung mittels Pre-Screening Referenzgruppe in Screening und Kontrolregionen) Alter, Follow-up, Zeitraum, Area	<i>Dilution/Kontamination</i> De-facto Screening ausserhalb des Programs wird als nicht Screening klassifiziert und dadurch kann die Effektivität von Screening unterschätzt werden
Jonsson 2000 ³⁰	Schweden	Kohorten	Multiplicative Poisson Regression (Referenzzeitraum als Kontrolle für sekuläre Trends)	k.A.	Follow-up Jahr, Zeitraum, Ort	<i>Dilution/Kontamination</i> De-facto Screening ausserhalb des Programs wird als nicht Screening klassifiziert und dadurch kann die Effektivität von Screening unterschätzt werden.
Jonsson 2003a ³¹	SW-Gav	Kohorten	Multiplicative Poisson Regression (Kovariaten waren Follow-up Zeit, County, Zeitraum)	Exzess Mortalität vs. Mortalität von BK als Todesursache.	Unterschiede in Hintergrund BK-Mortalität (Adjustierung mittels Pre-Screening Referenzgruppe in Screening und Kontrolregionen) Alter, Follow-up,	-
Jonsson 2003b ³²	Schweden	Kohorten	Multiplicative Poisson Regression (Kovariaten waren Follow-up Zeit, County, Zeitraum)	Exzess Mortalität vs. Mortalität von BK als Todesursache	Unterschiede in Hintergrund BK-Mortalität (Adjustierung mittels Pre-Screening Referenzgruppe in Screening und Kontrolregionen) Alter, Follow-up,	-
SOSSEG 2006 ^{33,34}	Schweden	Kohorten	Poisson Regression	Eingeladen/nicht eingeladen vs. gescreent/nicht gescreent.	Adjustierung für zeitlichen Trends in Inzidenz, Fatalität und Mortalität	Verbesserte Therapie
Tabar 2001 ^{35,36}	Two-County	Kohorten	Log-lineare Regression	Eingeladen/nicht eingeladen vs. gescreent/nicht gescreent. Gesamt mortalität, Gesamtkrebsmortalität, BK-Mortalität	Adjustierung für Veränderungen in BK-Inzidenz (siehe Seite 24). Alter	<i>Verbesserung der Therapie, Aufmerksamkeit für BK</i> Vergleich der BK-Mortalität in der verschiedenen Studienzeitraum bei Frauen die nie gescreent werden (20-39 J) um dies abzuschätzen (ca. 70% des Effektes wäre auf

Studie	Programm	Studien-Design	Statistische Verfahren	Sensitivitätsanalyse ^a	Confounder / Effekt Modifier	
					In der Analyse	In der Diskussion
						Screening zurückzuführen).
Ascunce 2007 ³⁸	Navarra	Kohorten	Joinpoint Regression Analysis (http://srab.cancer.gov/joinpoint/)	Mit oder ohne Prävalente Fälle	k.A.	k.A.
Quinn 1995 ⁴³	NHSBSP	Ökologische Studie	Vergleich der Altersstandardisierten BK-Mortalitätsraten	k.A.	Alter	Verbesserte Therapie
Barrichelli ¹³	Florenz	Ökologische Studie	Altersstandardisierten Raten Poisson-Regression Cox Regression	k.A.	Alter, Staging	Verbesserte Therapie
Gorini 2004 ¹⁴	Florenz	Ökologische Studie	Vergleich der altersstandardisierten BK-Mortalitätsraten in Gebiet mit früher (1970) Screening-Implementierung vs. Gebiete mit Screening seit 1990 Log-linear Regression Model	k.A.	Alter, Jahr des Todes	Verbesserte Therapie
Törnberg 1994 ³⁷	Schweden	Ökologische Studie	Multiplicatie Poisson Regression (Zusammenhang zwischen Extension des Mammographie-Programms und BK-Mortalität)	k.A.	Alter, Zeitraum, County	Verbesserung der Therapie über die Zeit
Hakama 1999 ^{9,10}	Cancer Society Finnland	quasi-randomisiert	Vergleich von standardisierter Mortalitätsrate (SMR) (standardisiert zu finnischer Bevölkerung)	Mit und ohne Todesfälle von inzidenten Fälle die vor ersten Screening diagnostiziert wurden	Alter (Standardisierte Mortalitätsrate)	k.A.
Hakama 1995 ¹¹	Kotka Pilot ^p	quasi-randomisiert	Vergleich von standardisierter Mortalitätsrate (SMR) (standardisiert zu finnischer Bevölkerung)	Mit und ohne Todesfälle von inzidenten Fälle die vor ersten Screening diagnostiziert wurden	Alterstandardisiert	Selbstuntersuchung, Education über BK (erhöhte Aufmerksamkeit) Untersuchung durch Krankenschwester
Thompson	Puget-Sound	Fall-Kohorte	Cox Regression	k.A.	Alter, familiäre Vorgeschichte,	

Studie	Programm	Studien-Design	Statistische Verfahren	Sensitivitätsanalyse ^a	Confounder / Effekt Modifier	
					<i>In der Analyse</i>	<i>In der Diskussion</i>
1994 ⁴⁴	Health Cooperative				Fertilität	

BK: Brustkrebs; CC: Fall-Kontrolle; GP: Hausarzt; HRT: Hormon-Ersatz-Therapie; k.A.: keine Angaben; NHSBSP-EA: National Health Service Breast Screening Programme East Anglia; SA: Southern Australia; SES: Sozioökonomisches Status; SMR: Standardisierte Mortalitätsrate

^aIn den Studien nicht explizit Sensitivitätsanalyse benannt.

5.3 Ergebnisse der Studien

Tabelle 14. Ergebnisse Brustkrebsmortalität.

Land	Studie	Design	Altersgruppe	Programm Effektivität ^a RR (95%-CI)	Mammographie-Scr. Effektivität ^b RR/OR (95%-CI)	
					Unkorrigiert	Korrigiert ^c
Australien	Roder 2008 ²	Fall-Kontroll	45-80 J	-	0,59 (0,47-0,74)	0,70 (k.A.)-
Grossbritannien	Fielder 2004 ⁴¹	Fall-Kontroll	50-75 J	-	0,62 (0,47-0,82)	0,75 (0,49-1,14)
	Moss 1992 ⁴²	Fall-Kontroll	45-64 J	0,76 (0,54-1,08) ^d	0,51 (0,27-0,98)	-
	Allgood 2005 ⁴⁰	Fall-Kontroll	50-70 J	0,65 (0,48-0,88)	0,35 (0,24-0,50)	0,52 (0,32-0,84)
Island	Gabe 2007 ¹²	Fall-Kontroll	40-69 J	-	0,59 (0,41-0,84)	0,65 (0,39-1,09)
Italien	Puliti 2008 ¹⁷	Fall-Kontroll	50-74 J	0,75 (0,62-0,92)	0,46 (0,38-0,56)	0,55 (0,36-0,85)
Niederlande	Collette 1984 ^{21,22,23,24}	Fall-Kontroll	50-64 J	-	0,54 (0,37-0,79)	0,64 (0,41-0,39) ^e
	van Dijck 1995 ²⁶	Fall-Kontroll	> 65 J	-	0,56 (0,28-1,13)	-
Dänemark	Olsen 2005 ^{5,6}	Kohorten	50-69 J	0,75 (0,63-0,89)	0,60 (0,49-0,74)	0,63 (k.A.)
Finnland	Anttila 2002 ⁷	Kohorten	50-59 J	0,81 (0,62-1,05)	-	-
	Parvinen 2006 ⁸	Kohorten	55-69 J	0,64 (0,47-0,88)	-	-
Italien	Paci 2002 ¹⁶	Kohorten	50-69 J	0,75 (0,54-1,04)	0,63 (0,42-0,94)	-
Kanada	Coldman 2006 ⁴	Kohorten	40-79 J	-	0,60 (0,55-0,65)	0,76 (k.A.)
Niederlande	Peer 1995 ²⁵	Kohorten	35-49 J	0,94 (0,68-1,29)	-	-
Schweden	Duffy 2002 ²⁷	Kohorten	40-69 J	0,68 (0,60-0,77)	0,56 (0,50-0,62)	0,61 (0,55-68)
	Jonsson 2007 ²⁸	Kohorten	40-74 J	0,70 (0,56-0,87)	-	-
	Jonsson 2001 ²⁹	Kohorten	50-69 J	0,84 (0,67-1,05)	-	-
	Jonsson 2000 ³⁰	Kohorten	40-49 J	0,91 (0,72-1,15)	-	-
	Jonsson 2003a ³¹	Kohorten	40-64 J	0,82 (0,65-1,03)	-	-
	Jonsson 2003b ³²	Kohorten	70-74 J	0,82 (0,57-1,19)	-	-
	SOSSEG ^{33,34}	Kohorten	40-69 J	0,73 (0,69-0,77)	0,55 (0,51-0,59)	0,57 (0,53-0,62)
	Tabar 2001 ³⁶	Kohorten	40-69 J	0,50 (0,41-0,60)	0,37 (0,30-0,46)	0,52 (0,43-0,63)
Spanien	Ascunce 2007 ³⁸	Kohorten	45-65 J	0,75 (0,63-0,89)	-	-
Schweden	Törnberg 1994 ³⁷	Ökologische Studie	50-74 J	0,81 (0,63-1,03)	-	-
Finnland	Hakama 1999 ^{9,10}	quasi-randomisiert	50-59 J	0,76 (0,53-1,09)	0,66 (k.A.)	-
	Hakama 1995 ¹¹	quasi-randomisiert	40-47 J	0,11 (0,00-0,71)	-	-
USA	Thompson 1994 ⁴⁴	Fall-Kohorte	> 40 J	-	0,80 (0,34-1,85)	-

k.A.: keine Angaben

^aVergleich zwischen Eingeladenen und Nicht Eingeladenen, Exponierte zu einen Screening-Programm vs. Nicht Exponierte (z.B. Pre-Screening Alterskohorte vs. Screeningalterskohorte, oder Region mit Screening vs. ohne Screening). ^bVergleich zwischen tatsächlich gescreente und nicht gescreente bzw. nicht eingeladen., ^cKorrigiert für Self-Selektion wenn nicht anders erläutert.

^dGescreent vs. eingeladen und nicht teilgenommen + nicht eingeladen. ^eFür Lead-Time-Bias korrigiert.

Folgende Studien berichten nicht Brustkrebsmortalität bzw. nicht in für diese Tabelle extrahierbarer Form: Barrichelli¹³, Broeders 2002²⁰, Gorini 2004¹⁴, Paci 2008¹⁵, Quinn 1995⁴³, Spillane 2001³

6. Literaturliste

6.1 Übermittelte Studien

6.1.1 Systematische Übersichten

1 Schopper D, de Wolf C. How effective are breast cancer screening programmes by mammography? Review of the current evidence. *European Journal of Cancer* 2009; 45: 1916-1923.

6.1.2 Primärliteratur

Australien

2 Roder D, Houssami N, Farshid G, Gill G, Luke C, Downey P, Beckmann K, Iosifidis P, Grieve L, Williamson I. Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia. *Breast Cancer Research and Treatment* 2008; 108: 409-416.

3 Spillane AJ, Kennedy CW, Gillett DJ, Carmalt HL, Janu NC, Rickard MT, Donnellan MJ. Screen-detected breast cancer compared to symptomatic presentation: an analysis of surgical treatment and end-points of effective mammographic screening. *Australia New Zealand Journal of Surgery* 2001; 71: 398-402.

Canada

4 Coldman A, Phillips N, Warren L, Kann L. Breast cancer mortality after screening mammography in British Columbia women. *International Journal of Cancer* 2006; 120: 1076-1080.

Dänemark

5 Olsen AH, Njor SH, Vejborg O, Schwartz W, Dalgaard P, Jensen MB, Brix-Tange U, Blichert-Toft M, Rank F, Mouridsen H, Lynge E. Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. *BMJ* 2005a; 330: 220-222 (abrd). <http://bmj.com/cgi/doi/10.1136/bmj.38313.639236.82>.

6 Olsen AH, Njor SH, Vejborg O, Schwartz W, Dalgaard P, Jensen MB, Brix-Tange U, Blichert-Toft M, Rank F, Mouridsen H, Lynge E. A model for determining the effect of mammography service screening. *Acta Oncologica* 2005b; 44: 120-128.

Finnland

7 Anttila A, Koskela J, Hakama M. Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. *Journal of Medical Screening* 2002; 9: 153-158.

8 Parvinen I, Helenius H, Pylkkänen L, Anttila A, Immonen-Räihä P, Kauhava L, Räsänen O, Klemi PJ. Service screening mammography reduces breast cancer mortality among elderly women in Turku. *Journal Medical Screening* 2006; 13: 34-40.

9 Hakama M, Pukkala E, Söderman B, Day N. Implementation of screening as a public health policy: issues in design and evaluation. *Journal of Medical Screening* 1999; 6: 209-216.

10 Hakama M, Pukkala E, Heikkilä M, Kallio M. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *British Medical Journal* 1997; 314: 864-867.

11 Hakama M, Pukkala E, Kallio M, Godenhjelm K, Svinhufvud U. Effectiveness of screening for breast cancer in women under 50 years at entry: the Kotka pilot project in Finland. *International Journal of Cancer* 1995; 63: 55-57.

Island

12 Gabe R, Tryggvadóttir L, Sigfússon BF, Ólafsdóttir GF, Sigurðsson K, Duffy SW. A case-control study to estimate the impact of the screening population-based mammography screening program on breast cancer death. *Acta Radiologica* 2007; 48: 948-955.

Italien

- 13 Barchielli A, Paci E. Trends in breast cancer mortality, incidence, and survival, and mammographic screening in Tuscany, Italy. *Cancer Causes and Control* 2001; 12: 249-255.
- 14 Gorini G, Zappa M, Miccinesi G, Paci E, Seniore Costantini A. Breast cancer mortality trends in two areas of the province of Florence, Italy, where screening programmes started in the 1970s and 1990s. *British Journal of Cancer* 2004; 90: 1780-1783.
- 15 Paci E, Coviello E, Miccinesi G, Puliti D, Cortesi L, de Lisi V, Ferretti S, Mangone L, Perlangeli V, Ponti A, Ravaioli A, Sassoli de' Bianchi P, Segnan N, Stracci F, Tumino R, Zarcone M, Zorzi M, Zappa M, IMPACT Working Group. Evaluation of service mammography screening impact in Italy: The contribution of hazard analysis. *European Journal of Cancer* 2008; 44: 858-865.
- 16 Paci E, Duffy SW, Giorgi D, Zappa M, Crocetti E, Vezzosi V, Bianchi S, Roselli del Turco M. Quantification of the effect of mammographic screening on fatal breast cancers: the Florence Programme 1990-96. *British Journal of Cancer* 2002; 87: 65-69.
- 17 Puliti D, Miccinesi G, Collina N, de Lisi V, Federico M, Ferretti S, Finarelli AC, Foca F, Mangone L, Naldoni C, Petrella M, Ponti A, Segnan N, Sigona A, Zarcone M, Zorzi M, Zappa M, Paci E, the IMPACT Working Group. Effectiveness of service screening: a case-control study to assess breast cancer mortality reduction. *British Journal of Cancer* 2008; 99: 423-427.

Japan

- 18 Kawai M, Kuriyama S, Suzuki A, Nishino Y, Ishida T, Ohnuki K, Amari M, Tsuji I, Ohuchi N. Effect of screening mammography on breast cancer survival in comparison to other detection methods: a retrospective cohort study. *Cancer Science* 2009; 100: 1479-1484.
- 19 Kuroishi T, Hirose K, Suzuki T, Tominaga S. Effectiveness of mass screening for breast cancer in Japan. *Breast Cancer* 2000; 7: 1-8.

Niederlande

- 20 Broeders MJM, Verbeek ALM, Straatman H, Peer PGM, Pasker-de Jong PCM, Beex LVAM, Hendriks JHCL, Holland R. Repeated mammographic screening reduces breast cancer mortality along the continuum of age. *Journal of Medical Screening* 2002; 9: 163-167.
- 21 Collette C, Collette HJ, Fracheboud J, Slotboom BJ, de Waard F. Evaluation of a breast cancer screening programme –the DOM project. *European Journal of Cancer* 1992; 28A: 1985-1988.
- 22 Collette HJ. Attempts to evaluate a non-randomized breast cancer screening programme (the 'DOM-project'). *Maturitas* 1985; 7: 43-50.
- 23 Collette HJ, Day NE, Rombach JJ, de Waard F. Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. *The Lancet* 1984; 1: 1224-1226.
- 24 Miltenburg GAJ, Peeters PHM, Fracheboud J, Collette HJA. Seventeen-year evaluation of breast cancer screening: the DOM project, The Netherlands. *British Journal of Cancer* 1998; 78: 962-965.
- 25 Peer PG, Werre JM, Mravunac M, Hendriks JH, Holland R, Verbeek AL. Effect on breast cancer mortality of biennial mammographic screening of women under age 50. *International Journal of Cancer* 1995; 60: 808-811.
- 26 van Dijk JAMM, Verbeek ALM, Beex LVAM, Hendriks JHCL, Holland R, Mravunac M, Straatman H, Werre JM. Mammographic screening after the age of 65 years: evidence for a reduction in breast cancer mortality. *International Journal of Cancer* 1996; 66: 727-731.

Schweden

- 27 Duffy SW, Tabar L, Chen HH, Holmqvist M, Yen MF, Abdalah S, Epstein B, Frodis W, Ljunberg E, Hedborg-Melander C, Sundbom A, Tholin M, Wiege M, Akerlund A, Wu HM, Tung TS, Chiu CP, Huang CC, Smith RA, Rosen M, Stenbeck M, Holmberg L. The impact of organized mammography service screening on breast carcinoma mortality in seven screened counties – a collaborative evaluation. *Cancer* 2002; 95:458-469.
- 28 Jonsson H, Bordás P, Wallin H, Nyström L, Lenner P. Service screening with mammography in Northern Sweden: effects on breast cancer mortality: an update. *Journal of Medical Screening* 2007; 14: 87-93.
- 29 Jonsson H, Nyström L, Törnberg S, Lenner P. Service screening with mammography of women aged 50-69 years in Sweden: effects on mortality from breast cancer. *Journal of Medical Screening* 2001; 8: 152-160.
- 30 Jonsson H, Törnberg S, Nyström L, Lenner P. Service screening with mammography in Sweden. Evaluation of effects of screening on breast cancer mortality in age group 40-49 years. *Acta Oncologica* 2000; 39: 617-623.
- 31 Jonsson H, Nyström L, Törnberg S, Lundgren B, Lenner P. Service screening with mammography. Long-term effects on breast cancer mortality in the county of Gävleborg, Sweden. *The Breast* 2003a; 12: 183-193.
- 32 Jonsson H, Törnberg S, Nyström L, Lenner P. Service screening with mammography of women aged 70-74 years in Sweden. Effects on breast cancer mortality. *Cancer Detection and Prevention* 2003b; 27: 360-369.

- 33 Swedish Organised Service Screening Evaluation Group. Reduction in Breast Cancer Mortality from Organized Service Screening with Mammography: 1. Further Confirmation with Extended Data. *Cancer Epidemiol Biomarkers Prev* 2006; 15(1): 45-51.
- 34 Swedish Organised Service Screening Evaluation Group Reduction in Breast Cancer Mortality from the Organised Service Screening with Mammography: 2. Validation with Alternative Analytic Methods. *Cancer Epidemiol Biomarkers Prev* 2006; 15(1): 52-56.
- 35 Tabar L, Ming-Fang Y, Bedrich V, Hsiu-I TC, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet* 2003; 361: 1405-1410.
- 36 Tabar L, Bedrich V, Ming-Fang Y, Hsiu-I TC, Duffy SW, Smith RA. Beyond Randomized Controlled Trials Organized Mammographic Screening Substantially Reduces Breast Carcinoma Mortality. *Cancer* 2001; 91: 1724-1731.
- 37 Törnberg S, Carstensen J, Hakulinen T, Lenner P, Hatschek T, Lundgren B. Evaluation of the effect on breast cancer mortality of population based mammography screening programmes. *Journal of Medical Screening* 1994; 1: 184-7.

Spanien

- 38 Ascunce EN, Moreno-Iribas M, Barcos Urriaga A, Ardanaz E, Ederra Sanz M, Castilla J, Egües N. Changes in breast cancer mortality in Navarre (Spain) after introduction of a screening programme. *Journal of Medical Screening* 2007; 14: 14-20.
- 39 Bonfill Cosp X, Marzo Castillejo M, Sentís Crivillé M, Rossell Mir R, Gallardo Cistaré X, Florensa Masip R, Rivero Ferrer E, Moreno Quiroga C. Evaluation of the regular practice of breast cancer screening in a health area. *International Journal of Technology Assessment in Health Care* 1996; 12: 388-94.

UK

- 40 Allgood PC, Warwick J, Warren RML, Day NE, Duffy SW: A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *British Journal of Cancer* 2008; 98: 206-209.
- 41 Fielder HM, Warwick J, Brook D, Gower-Thomas K, Cuzick J, Monypenny I, Duffy SW. A case-control study to estimate the impact on breast cancer death of the breast screening programme in Wales. *Journal of Medical Screening* 2004; 11: 194-198.
- 42 Moss SM, Summerley ME, Thomas BT, Ellman R, Chamberlain JO. A case-control evaluation of the effect of breast cancer screening in the United Kingdom trial of early detection of breast cancer. *Journal of Epidemiology and Community Health* 1992; 46: 362-364.
- 43 Quinn M, Allen E on behalf of the UK Association of Cancer Registries. Changes in incidence of and mortality from breast cancer in England and Wales. *British Medical Journal* 1995; 311: 1391-1395.

USA

- 44 Thompson RS, Barlow WE, Taplin SH, Grothaus L, Immanuel V, Salazar A, Wagner EH. A population-based case-cohort evaluation of the efficacy of mammographic screening for breast cancer. *American Journal of Epidemiology* 1994; 140: 889-901.

6.2 Weitere Literatur

- 45 Biesheuvel C, Barrat A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: a systematic review. *Lancet Oncology* 2007; 8: 1129-1138.
- 46 Cronin KA, Weed DL, Connor RJ, Prorok PC. Case-control studies of cancer screening: theory and practice. *Journal of the National Cancer Institute* 1998; 90:498-504.
- 47 Duffy SW, Nagtegaal ID, Wallis M, Cafferty FH, Houssami N, Warwick J, Allgood PC, Kearins O, Tappenden N, O'Sullivan E, Lawrence G. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *American Journal of Epidemiology* 2008; 168: 98-104.
- 48 Duffy S, Cuzick J, Tabar L, Vitak B, Chen H, Yen M, Smith R. Correcting for non-compliance bias in case-control studies to evaluate cancer screening programmes. *Journal of the Royal Statistical Society Series C* 2002; 51:235-243.
- 49 Duffy SW, Olsen AH, Gabe R, Tabar L, Warwick J, Fielder H, Tryggvadottir, Agbaje OF. Screening opportunity bias in case-control studies of Cancer Screening. *Journal of Applied Statistics* 2008; 35: 537-546.
- 50 Hennekens CH, Buring J. *Epidemiology in Medicine*. Boston – Toronto: 1987.
- 51 Lenner P, Jonsson H. Excess mortality from breast cancer in relation to mammography in Sweden. *Journal of Medical Screening* 1997; 4: 6-9.
- 52 Miller AB, Goel V. Screening. In: Detels R, McEwen J, Beaglehoe R, Tanaka H (Hrsg.). *Oxford Textbook of Public Health* (4th Edition). Oxford – New York 2002. S. 1823-1838.

- 53 Morrison AS. Screening. In: Rothman KJ, Greenland S. *Modern Epidemiology* (2nd Edition). Philadelphia 1998. S 499-518.
- 54 Olsen AH, Bihrmann K, Jensen M-B, Vejborg I, Lynge E. Breast density and outcome of mammography screening: a cohort-study. *British Journal of Cancer* 2009; 100: 1205-1208.
- 55 Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 2001; 358: 1340-1342.
- 56 Rettig B, Horton M. Breast cancer in Nebraska: recent trends in incidence, mortality, and screening. *Nebraska Medical Journal* 1996; 81: 63-9.
- 57 Walter U, Schwartz FW, Robra BP, Schmidt T. Prävention. In :Schwartz FW, Badura B, Busse R, Raspe H, Siegrist J, Walter U. *Das Public Health Buch: Gesundheit und Gesundheitswesen* (2. Auflage). München – Jena 2003. S. 189-214.

7. Anhang

Recherche Strategie

Search	Most Recent Queries	Result
#1	Search "Mammography" [MeSH]	19364
#2	Search "Mass Screening" [MeSH]	94699
#3	Search #1 OR #2	108964
#4	Search "Breast Neoplasms/epidemiology" [MeSH]	23862
#5	Search "Breast Neoplasms" [MeSH]	164992
#6	Search #4 OR #5	164992
#7	Search #3 AND #6	18793
#8	Search (Breast cancer AND screening AND evaluat*[ti])	2835
#9	Search (Breast cancer AND screening AND effect*[ti])	3999
#10	Search #8 OR #9	6741
#11	Search #7 AND #10	1259

Systematischer Review von Schopper & Wolf

Eine der 44 übermittelten Publikationen war eine systematische Übersichtsarbeit über die Effektivität von Mammographie-Screening-Programmen in verschiedenen Ländern (1). Diese Arbeit fasst die Ergebnisse zu Screeningprogramme in Australien, Canada, Dänemark, Finnland, Großbritannien, Island, Italien, den Niederlanden, Schweden und Spanien. Insgesamt umfasst die Arbeit 20 Publikationen zu den verschiedenen Programmen, wie in Tabelle 15 gezeigt.

Tabelle 15. Systematische Übersichtsarbeit Schooper & Wolf 2009.

Land	Region	Publikationen	Liegt vor?
Australien		AIHW 2006	NEIN
		Roder 2008	JA
Canada	British Columbia	Coldman 2007	JA
Dänemark	Kopenhagen	Olsen 2005	JA
Finnland	National	Sarkeala 2008	NEIN
	Helsinki	Anttila 2002	JA
	Turku	Parvinen 2006	JA
Island	national	Gabe 2007	JA
Italien	national	Paci 2008	JA
Niederlande	national	Otto 2003	NEIN
		Erasmus 2006	NEIN
		Fracheboud 2004	NEIN
Schweden	9-counties	SOSSEG 2006a	JA
		SOSSEG 2006b	JA
	2-counties	Tabar 2001	JA
	national	Jonsson 2001	JA
	Northern Sweden	Jonsson 2007	JA
Spanien	Navarra	Ascunce 2007	JA
UK	East Anglia	Allgood 2008	JA
	National	Cancer Research 2009	NEIN
AIHW: Australian Institute of Health and Welfare; SOSSEG: Swedish Organised Service Screening Evaluation Group			